# Journal of Mathematics and Physics

(Founded by C. L. E. Moore)

## Volume XXXIII

1954 -55

*Board of Editors*

E. REISSNER, *Managing Editor*

H. FESHBACH
P. FRANKLIN
E. A. GUILLEMIN
C. C. LIN
P. M. MORSE
D. J. STRUIK
N. WIENER

*The Technology Press*
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

QA 1
J87
v.33.

## CONTENTS


Page

Asymptotic Solutions of the Stability Equations of a Compressible Fluid
*By Cathleen S. Morawetz* ..... 1

Comparison Theorems for Subsonic Flows
*By J. B. Serrin* ..... 27

The Effect of Thickness on Pitching Airfoils at Supersonic Speeds
*By John C. Martin and Nathan Gerber* ..... 46

Lattice Theoretic Properties of Frontal Switching Functions
*By E. N. Gilbert* ..... 57

A General Solution of the Two-Frequency Modulation Product Problem II. Tables of the Functions $A_{mn}(h, k)$
*By Robert L. Sternberg* ..... 68

On the Degree of Convergence of Solutions of Difference Equations to the Solution of the Dirichlet Problem
*By J. L. Walsh and David Young* ..... 80

The Loaded Virbrating Net and Resulting Boundary-Value Problems for a Partial Difference Equation of the Second Order
*By Tomlinson Fort* ..... 94

The Convergence of the Solutions of a Class of Iterative Difference Equations
*By J. Samoloff* ..... 105

Regular and Irregular Coulomb Wave Functions Expressed in Terms of Bessel-Clifford Functions
*By Milton Abramowitz* ..... 111

On a Perturbation Theory Based on the Method of Characteristics
*By C. C. Lin* ..... 117

Linearization of the Equations of Non-Steady Flow in a Compressible Fluid
*By John W. Miles* ..... 135

Stress and Velocity Fields in Soil Mechanics
*By R. T. Shield* ..... 144

Loading Paths and the Incremental Strain Law
*By G. H. Handelman and W. H. Warner* ..... 157

The Operational Calculus of Legendre Transforms
*By R. V. Churchill* ..... 165

Asymptotic Properties of Bessel Functions and the Radiation from a Synchrotron
*By David Park* ..... 179

On Electrical Network Determinants
*By N. F. Tsang* ..... 185

Conversion of a Brune Cycle With an Ideal Transformer into a Cycle Without an Ideal Transformer
*By F. M. Reza* ..... 194

A General Solution of the Two-Frequency Modulation Product Problem III. Rectifiers and Limiters
*By Robert L. Sternberg* ..... 199

A Theory for Hydrofoils of Finite Span
*By Y. T. Wu* ..... 207

A Note on the Asymptotic Solutions of the Equation of Hydrodynamic Stability
*By Richard C. Di Prima* ..... 249

Coding to Achieve Markov Type Redundancy
*By Joseph E. Flanagan* ..... 258

An Application of Sommerfeld's Complex Order Wave Functions to Antenna Theory
*By Charles H. Papas* ..... 269




A 89825

The Orthogonality of the Hyperboloid Functions
*By Robert G. Van Nostrand* ..... 276

On the Calculation of Integrals of the Form $\int_0^{\theta} \sin^n\theta \cos^n\theta \, d\theta$
*By Henry E. Fettis* ..... 283

New Bounds in Harmonic and Biharmonic Problems
*By L. E. Payne and H. F. Weinberger* ..... 291

On the Dynamics of Small Vapor Bubbles in Liquids
*By S. A. Zwick and M. S. Plesset* ..... 308

A General Theory of Wideband Matching with Dissipative 4 Poles
*By R. La Rosa and Herbert J. Carlin* ..... 331

The Solution by Iteration of Nonlinear Integral Equations
*By Mark Lotkin* ..... 346

A Simple Method for Summing Certain Slowly Convergent Series
*By Herbert E. Salzer* ..... 356

Tables of Self-Equilibrating Functions
*By G. Horvay and J. S. Born* ..... 360

Note on an Integral of Ramanujan
*By Malcolm K. Brachman* ..... 374

# ASYMPTOTIC SOLUTIONS OF THE STABILITY EQUATIONS OF A COMPRESSIBLE FLUID

By Cathleen S. Morawetz

**1. Introduction.** In the study of the stability of a compressible fluid carried out by Lees and Lin [1], use is made of formal asymptotic series which are formal solutions of a sixth order system of differential equations involving a large parameter. The principal feature of this system is that if we let the parameter go to infinity the system is reduced to one of second order with a singular point. In this paper* we shall prove that these formal series actually represent solutions of the given system, and investigate some of the properties of these solutions. In many problems of this type, it is sufficient to know that there exists some domain of validity for the formal series solutions. Here, however, our principal object will be to find a sufficiently large domain of validity so that the boundary value problem arising from the stability problem can be treated asymptotically. For this purpose, we need a large domain of validity for a complete set of asymptotic solutions.

In addition, we are interested in finding out how the solutions behave as we proceed around the singular point of the reduced system.

General theorems for systems of arbitrary order with a real variable and a complex parameter have been proved by H. Turrittin [2] who has established the existence of formal asymptotic solutions and determined sufficient conditions on their domains of validity. Similar results have been obtained by M. Hukuhara [3] for complex variable and complex parameter. These domains are, however, too small for our purposes and cannot yield the "Stokes' phenomena" that are associated with the singular point. Methods similar to those used here for an analogous fourth order differential equation were developed by Wasow [4]. Certain complications arise because the system considered is of higher order. We need the equivalent of "factoring" a differential equation and we have many more separate domains to consider in the complex domain of the independent variable. The general results are very similar; in particular, the existence is established of solutions which behave like power series in the parameter in one region and diverge exponentially in another.

The investigation of these asymptotic solutions leads to the following two conclusions of practical interest.

(1) *It is proved rigorously that the proper branch of the multiple-valued asymptotic solutions has been used in* [1].

(2) *The existence or absence of an inner viscous region and its nature are rigorously established.* This corroborates and extends the results in [1]. Briefly, the conclusions are as follows:

(a) In the case of *unstable* disturbances with a finite rate of amplification, there is *no inner viscous region* if the Reynolds number of the mean flow is large enough.

* The critical review of this paper by Dr. H. Turrittin is gratefully acknowledged and very much appreciated.

(b) In the case of *stable* disturbances with finite damping, there is a *finite viscous region* in the interior of the fluid, no matter how large the Reynolds number becomes. In fact, a minimum width for this region is established.

(c) In the case of neutral disturbances, other methods of analysis indicate that there must be an inner viscous layer whose thickness approaches zero as the Reynolds number becomes infinite. The methods used in the present paper cannot establish the existence of such a layer. However, it is definitely proved in the case of constant Prandtl number that if a viscous region does exist, its width is not finite, but shrinks to zero as the Reynolds number becomes infinite.

The conclusions in cases (a) and (c) are exactly the same as in [1]. In case (b), the earlier method of study enabled the authors to recognize only certain viscous *layers* associated with the viscous *region*. No conclusion was drawn about the behavior of the fluid between such layers. The present investigation, therefore, yields more complete information about the influence of viscosity in the case of damped disturbances.

An immediate deduction from these results (whether in the present form or in the less complete form in [1]) is that *damped disturbances and unstable disturbances are not complex conjugates in the limiting case of vanishing viscosity*, although such a conclusion would be reached very easily by formally examining the differential equations satisfied by such disturbances.

The present results are similar to those obtained by Wasow [4], who corroborated and extended the earlier result of Lin [5] in the incompressible case.

It should be remarked that all of the above conclusions hold, provided the boundary of the flow is not too "close" to the viscous region discussed above, that is, provided the distance between the viscous region and the boundary remains finite as the Reynolds number becomes infinite.

The general case corresponding to unstable or stable disturbances is contained in Sections 6 and 7, Theorems 1 to 4. The case of constant Prandtl number is in Section 8, Theorem 5. Neutral disturbances correspond to a special case of the mathematical problem which cannot be handled by the direct methods used here unless the Prandtl number is constant. The existence of the inner viscous region is contained in the results of Section 9. The system for which these theorems are proved is a slight generalization of the system which arises in the stability of parallel flow. In Section 2, the leading terms in the asymptotic expansions for the flow quantities are determined. A general discussion of the hydrodynamic problem is contained in Section 10.

This investigation was sponsored by and conducted principally with the financial assistance of the National Advisory Committee for Aeronautics at the Massachusetts Institute of Technology. Some revisions were carried out under Navy contract Nonr-285(02) at the Institute for Mathematical Sciences, New York University.

**2. Differential Equations and Formal Solutions.** The differential equations for the amplitudes of small disturbances in parallel or nearly parallel main flow

are:

$$\alpha\rho\{i(w-c)f+w'\phi\} = -i\alpha\pi/\gamma M^2 + (\mu_1/R)\{f''+\alpha^2(i\phi'-2f)\} + \tfrac{2}{3}\alpha^2[(\mu_2-\mu_1)/R](-f+i\phi') + (1/R)\{m_1w''+m_1'w'+\mu_1'(f'+i\alpha^2\phi)\} \tag{2.1}$$

$$\alpha^2\rho\{i(w-c)\phi\} = -\pi'/\gamma M^2 - r/F^2 + (\mu_1\alpha/R)\{2\phi''-if'-\alpha^2\phi\} + \tfrac{2}{3}\alpha[(\mu_2-\mu_1)/R]\{\phi''+if'\} + (\alpha/R)\{im_1w' + 2\mu_1'\phi' + \tfrac{2}{3}(\mu_2-\mu_1)(\phi'+if)\} \tag{2.2}$$

$$i(w-c)r + \rho(\phi'+if) + \rho'\phi = 0 \tag{2.3}$$

$$\alpha\rho\{i(w-c)\theta + T'\phi\} = -\alpha(\gamma-1)\rho T(\phi'+if) + (\gamma/R\sigma_0)\{\mu_1(\theta''-\alpha^2\theta) + (m_1T')' + \mu_1'\theta'\} + R^{-1}\gamma(\gamma-1)M^2\{m_1w'^2 + 2\mu_1w'(f'+i\alpha^2\phi)\} \tag{2.4}$$

$$\pi/p = (r/\rho) + (\theta/T). \tag{2.5}$$

The notation is the same as that of [1]. The dependent variables are $f$, $\phi$, $r$, $\theta$, $\pi$. The independent variable is $y$, the distance perpendicular to the main flow. The large parameter is the Reynolds number $R$. The remaining quantities are known functions of $y$ or constants. The function $m_1$ is related to $\theta$ by $m_1 = (d\mu_1/dT)\theta$. We reduce these equations to six differential equations each of first order by setting, as in [1],

$$\begin{aligned} Z_1 &= f & Z_3 &= \phi & Z_5 &= \theta \\ Z_2 &= f' & Z_4 &= \pi/M^2 & Z_6 &= \theta'. \end{aligned} \tag{2.6}$$

Thus at once

$$Z_1' = Z_2 \tag{2.7}$$

$$Z_5' = Z_6 \tag{2.8}$$

and from (2.3) and (2.5)

$$Z_3' = -iZ_1 - (\rho'/\rho)Z_3 + i(w-c)[T^{-1}Z_5 - (M^2/p)Z_4]. \tag{2.9}$$

From (2.5), (2.3) and (2.1),

$$\begin{aligned} Z_2' = {} & \frac{\alpha R}{\mu_1}\left\{\frac{i}{\gamma}Z_4 + \rho i(w-c)Z_1 + \rho w'Z_3\right\} + \alpha^2 Z_1 \\ & + Z_3\left\{\frac{i\rho'\alpha^2}{\rho}\left[1+\tfrac{2}{3}\frac{\mu_2-\mu_1}{\mu_1}\right] - \frac{i\alpha^2\mu_1'}{\mu_1}\right\} \\ & - \frac{M^2\alpha^2}{p}Z_4(w-c)\left[1+\tfrac{2}{3}\frac{\mu_2-\mu_1}{\mu_1}\right] \\ & + \frac{1}{T}Z_5\left\{\alpha^2(w-c)\left[1+\tfrac{2}{3}\frac{\mu_2-\mu_1}{\mu_1}\right] - \frac{\eta_1 T}{\mu_1}\right\} - \frac{\eta_2}{\mu_1}Z_6 - \frac{\mu_1'}{\mu_1}Z_2. \end{aligned} \tag{2.10}$$

Here

$$\eta_1 = [(d\bar{\mu}_1/dT)w']', \qquad \eta_2 = (d\bar{\mu}_1/dT)w'.$$

From (2.5), (2.6) and (2.4)

$$\begin{aligned} Z_6' = {} & \frac{\alpha R\sigma_0}{\bar{\mu}_1}\left\{i(w-c)\rho Z_5 + \rho T' Z_3 - \frac{\gamma-1}{\gamma}[p'Z_3 + i(w-c)M^2 Z_4]\right\} \\ & + Z_5\left\{\alpha^2 - \eta_3 - \frac{\sigma_0(\gamma-1)}{\bar{\mu}_1}M^2\eta_5\right\} + Z_6\left[-\eta_4 - \frac{\bar{\mu}_1'}{\bar{\mu}_1}\right] \\ & + Z_2\{-2w'\sigma_0(\gamma-1)M^2\} + Z_3\{-2i\alpha^2 w'\sigma_0(\gamma-1)M^2\}. \end{aligned} \tag{2.11}$$

Here

$$\eta_3 = \frac{1}{\bar{\mu}_1}\left(T'\frac{d\bar{\mu}_1}{dT}\right)', \qquad \eta_4 = \frac{1}{\bar{\mu}_1}T'\frac{d\bar{\mu}_1}{dT}, \qquad \eta_5 = \frac{d\bar{\mu}_1}{dT}.$$

After some calculation we also find from (2.2)

$$\begin{aligned} Z_4' = {} & \{\gamma^{-1} + \tfrac{2}{3}\tfrac{1}{3}M^2 i\alpha R^{-1}p^{-1}(\bar{\mu}_2 + 2\bar{\mu}_1)(w-c)\}^{-1} \\ & \left\{-\frac{1}{F^2}\left(\frac{\rho}{p}M^2 Z_4 - \frac{\rho}{T}Z_5\right) - i\alpha^2(w-c)\rho Z_5 + \frac{1}{\alpha R}0(1)\right\} \end{aligned} \tag{2.12}$$

where $0(1)$ represents terms independent of $\alpha R$. Note that in (2.10) and (2.11) we have calculated higher terms than are given in [1], p. 18.

There are six formal asymptotic solutions to these equations. The first four solutions $Z_{ip}$ of (2.7)–(2.12) are

$$Z_{ip} \sim e^{\lambda Q_p}\sum_{\nu=0}^{\infty}\frac{f_{ip\nu}}{\lambda^\nu} \qquad i = 1, 2, \cdots, 6; p = 1, 2, 3, 4 \tag{2.13}$$

and the other two are

$$Z_{ip} \sim \sum_{\nu=0}^{\infty}\frac{f_{ip\nu}}{\lambda^\nu} \qquad i = 1, 2, \cdots 6; p = 5, 6 \tag{2.14}$$

with

$$\lambda = (\alpha R)^{\frac{1}{3}}.$$

By substituting (2.14) in (2.7)–(2.12) we find for $p = 5, 6$:

$$\begin{aligned} f_{1p0}' &= f_{2p0} \qquad f_{5p0}' = f_{6p0} \\ f_{2p0}' &= -if_{1p0} - (\rho'/\rho)f_{3p0} + i(w-c)[T^{-1}f_{5p0} - (M^2/p)f_{4p0}] \\ 0 &= (i/\gamma)f_{4p0} + \rho i(w-c)f_{1p0} + \rho w' f_{3p0} \\ 0 &= i(w-c)\rho f_{5p0} + \rho T' f_{3p0} - \frac{\gamma-1}{\gamma}[p'f_{3p0} + i(w-c)M^2 f_{4p0}] \\ f_{4p0}' &= \gamma\left[-\frac{1}{F^2}\left(\frac{\rho}{p}M^2 f_{4p0} - \frac{\rho}{T}f_{5p0}\right) - i\alpha^2(w-c)\rho f_{3p0}\right]. \end{aligned} \tag{2.15}$$

The system (2.15) can be reduced to a second order system of differential equations. They are the equations for inviscid flow which we would obtain from (2.7)–(2.12) by letting $\lambda^2 \to \infty$. We may select any two solutions for the leading terms of (2.13). Higher terms are obtained by quadratures.

As in the case of incompressible flow, one of the solutions of the inviscid equations is regular and the other solution has a logarithmic singularity at $w(y) = c$, see [1], p. 43.

By substituting (2.13) in (2.7)–(2.12) we find immediately from the highest coefficients in $\lambda$ in each equation:

$$f_{1p0} = f_{5p0} = f_{3p0} = f_{4p0} = 0 \qquad p = 1, 2, 3, 4$$

and from the next coefficient

$$Q_1 = -Q_2 = \int_{y_c}^{y} \sqrt{(i/\nu_1)(w - c)}\, dy$$

(2.16)

$$Q_3 = -Q_4 = \int_{y_c}^{y} \sqrt{(i\sigma_0/\nu_1)(w - c)}\, dy$$

and for $p = 1, 2$,

$$f_{5p1} = f_{3p1} = f_{6p0} = f_{4p1} = 0, \qquad Q_p' f_{1p1} = f_{2p0} \tag{2.17}$$

while for $p = 3, 4$,

$$f_{1p1} = f_{2p0} = f_{3p1} = f_{4p1} = 0, \qquad Q_p' f_{5p1} = f_{6p0}\,. \tag{2.18}$$

From the next coefficient in $\lambda$ of (2.13) we finally obtain, using (2.16) and (2.17),

$$2f_{1p1}' + f_{1p1}\left[\frac{Q_p''}{Q_p'} + \frac{i\rho w'}{(Q_p')^2\bar{\mu}_1} + \frac{\bar{\mu}_1'}{\bar{\mu}_1}\right] = 0.$$

Thus we may take, for $p = 1, 2$

$$\bar{\mu}_1^{-\frac{1}{2}}[Q_p'(w - c)]^{-\frac{1}{4}}$$

and

$$f_{2p0} = Q_p' f_{1p1}$$

Similarly we obtain from (2.13), (2.16) and (2.18) for $p = 3, 4$,

$$f_{6p1} = \frac{f_{6p0}}{Q_p'},$$

$$f_{6p0} = \frac{1}{\bar{\mu}_1\sqrt{Q_p'}}\, T^{\frac{1}{2}} p^{-(\gamma-1/2\gamma)} \exp\left(-\frac{\gamma - 1}{2\gamma}\frac{M^2}{F^2}(y - y_c) + \tfrac{1}{2}\int_{y_c}^{y} \eta_1\, dy\right).$$

The higher coefficients $f_{jp\nu}$ are obtained successively by quadratures.

The present purpose is to find where the expressions (2.13) and (2.14) represent solutions of (2.7)–(2.12). The results which are of general interest in the hydrodynamic problem are contained in Section 10. They follow from theorems given in Sections 3–9 which are proved for a more general system than (2.7)–(2.12).

**3. A More General Sixth Order System.** We shall prove our results for a system of more general form than (2.7)–(2.12) which retains its most important features. Consider a system of six differential equations, each of first order:

$$\frac{dz_k}{dx} = \sum_{p=1}^{6} \gamma_{kp}(x, \lambda) z_p \qquad k = 1, 2, \cdots, 6 \tag{3.1}$$

where $\gamma_{kp}$ are regular analytic functions of $x$ in a neighborhood $S$ of $x = 0$ and for sufficiently large values of $|\lambda|$. All values of $\lambda$ to be considered are on the same ray of the complex $\lambda$-plane. The coefficients $\gamma_{kl}$ are described for sufficiently large $|\lambda|$ by

$$\gamma_{kl}(x, \lambda) = \lambda^2 \left[ \sum_{\nu=0}^{\nu_1} \gamma_{kl\nu}(x) \lambda^{-\nu} + \frac{E(S)}{\lambda^{\nu_1+1}} \right]. \tag{3.2}$$

Here and elsewhere the generic symbol $E(S)$ denotes any function of $x$ and $\lambda$ which together with all its derivatives with respect to $x$ is bounded uniformly in $\lambda$ in every closed subdomain of $S$. This system has six formal solutions which are assumed to be of the form

$$e^{\lambda Q_p(x)} \sum_{\nu=0}^{\infty} \frac{f_{jp\nu}(x)}{\lambda^\nu} \qquad \begin{matrix} p = 1, \cdots, 6 \\ j = 1, \cdots, 6 \end{matrix} \tag{3.3}$$

with

$$\begin{aligned} Q_1 &= -Q_2 = Q = \int_0^x \sqrt{b(t)}\, dt \\ Q_3 &= -Q_4 = R = \int_0^x \sqrt{a(t)}\, dt \\ Q_5 &= Q_6 = 0. \end{aligned} \tag{3.4}$$

These solutions are formally independent in the sense that the formal expansion of their Wronskian does not vanish identically.

We assume $a(0) = b(0) = 0$ and that $a(x)$, $b(x)$ do not vanish anywhere else in $S$. In addition $|a'(0)| \neq |b'(0)|$.

The values of arg $a'(0)$ and arg $b'(0)$ play an important role. Case I, arg $a'(0) \neq$ arg $b'(0)$ will be treated in general. Case II, arg $a'(0) =$ arg $b'(0)$ will be treated only for arg $a'(0) \equiv$ arg $b'(0)$. This case occurs wherever we have neutral solutions of the stability equations and the Prandtl number is constant.

We shall prove the validity of these asymptotic expansions in certain subdomains of a doubly-connected domain $S$ surrounding the origin. The actual boundaries of $S$ will be defined later.

**4. Trjitzinsky's Lemma and the Reduction of a System of $n^{\text{th}}$ Order.** There are two tools which we use to show that finite sums of the expressions (3.3) actually represent solutions of (3.1) asymptotically. The first is Trjitzinsky's lemma, [6], p. 24 as extended to complex variable, see [4]. This lemma must be

used every time we prove that any one of (3.3) represents a solution. The second tool is the reduction of a system of order $n$ to a system of order $n-2$ using two known solutions of the $n$th order system in order to apply Trjitzinsky's lemma again. We formulate these two fundamental lemmas as we shall use them.

LEMMA 1. Let

$$\sum_{\nu=\nu_0}^{\infty} c_{jk,\nu}(x)\lambda^{-\nu} \qquad j = 1, \cdots, n, \quad k = 1, \cdots, n \tag{4.0}$$

be $n^2$ not necessarily convergent series with coefficients that are regular analytic in a domain $T$ and let $c_{jk}$ be any functions of the form

$$c_{jk}(x, \lambda) = \sum_{\nu=\nu_0}^{\nu_1} c_{jk,\nu}(x)\lambda^{-\nu} + E(T)\lambda^{-\nu_1-1}. \tag{4.1}$$

Now consider the system of differential equations

$$\frac{dy_j}{dx} = \sum_{k=1}^{n} c_{jk}(x, \lambda)y_k. \tag{4.2}$$

Suppose there exist $n^2$ series of the form

$$e^{\lambda P_p(x)} \sum_{\nu=0}^{\infty} f_{jp\nu}(x)\lambda^{-\nu} \qquad j = 1, \cdots, n \qquad p = 1, \cdots, n \tag{4.3}$$

which formally satisfy for each $p$ the formal differential equations obtained by replacing $c_{jk}(x, \lambda)$ in (4.2) by the series (4.0).

Let $T^*$ be a subdomain of $T$, every point of which can be reached from a fixed point $P$ by a path $\Gamma$ along which the $\tau n$ quantities

$$\operatorname{Re}(P_i - P_j) \qquad i = 1, 2, \cdots, \tau, \qquad j = 1, 2, \cdots, n$$

are monotonically non-decreasing. Then there exist $\tau$ linearly independent solutions

$$y_j = y_{jr}(x, \lambda) \qquad r = 1, 2, \cdots, \tau$$

of (4.2) for which

$$y_{jr}(x, \lambda) = e^{\lambda P_r(x)} [\sum_{\nu=0}^{\nu_2} f_{jr\nu}(x)\lambda^{-\nu} + E(T^*)\lambda^{-\nu_2-1}]. \tag{4.4}$$

LEMMA 2. If $x_{k1}$ and $x_{k2}$ are two independent solutions of the $n^{\text{th}}$ order system (4.2), a set of $n$ linearly independent solutions of (4.2) is given by $x_{k1}$, $x_{k2}$ and

$$\begin{aligned} x_{1r} &= x_{11}p_{1r} + x_{12}p_{2r} && r = 3, \cdots, n \\ x_{2r} &= x_{21}p_{1r} + x_{22}p_{2r} && r = 3, \cdots, n \\ x_{jr} &= x_{j1}p_{1r} + x_{j2}p_{2r} + z_{jr} && j = 3, \cdots, n \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} p_{1r} &= \sum_{p=3}^{n} \int_{\omega_{1r}}^{x} \frac{z_{pr}[c_{1p}x_{22} - c_{2p}x_{12}]}{W_{12}}\, dx \\ p_{2r} &= -\sum_{p=3}^{n} \int_{\omega_{2r}}^{x} \frac{z_{pr}[c_{1p}x_{21} - c_{2p}x_{11}]}{W_{12}}\, dx \end{aligned} \tag{4.6}$$

with

$$W_{jp} = \begin{vmatrix} x_{j1} & x_{j2} \\ x_{p1} & x_{p2} \end{vmatrix}; \tag{4.7}$$

$\omega_{1r}$ and $\omega_{2r}$ are arbitrary constants, $z_{jr}$, $j = 3, \cdots, n$, $r = 3, \cdots, n$ are $n - 2$ independent solutions of

$$\frac{dz_j}{dx} = \sum_{p=3}^{n} b_{jp} z_p, \qquad b_{jp} = c_{jp} = \frac{W_{j2}}{W_{12}} c_{1p} - \frac{W_{1j}}{W_{12}} c_{2p} \qquad \begin{matrix} j = 3, 4, \cdots, n \\ p = 3, 4, \cdots, n. \end{matrix} \tag{4.8}$$

The solutions (4.5) exist where $W_{12} \neq 0$ and remain independent of the path of integration if $W_{12} = 0$ at isolated points. Note that $W_{12} \not\equiv 0$ since $x_{k1}$ and $x_{k2}$ are independent.

Lemma 2 follows from the step-by-step reduction of a system of equations using one known solution at each step, see for example [7], p. 174. Using this procedure we find that the transformation

$$y_j = z_j + \frac{W_{j2}}{W_{12}} y_1 + \frac{W_{1j}}{W_{12}} y_2 \qquad j = 3, \cdots, n \tag{4.9}$$

which is valid for $W_{12} \neq 0$ reduces the differential equations (4.2) to $z_1 \equiv z_2 \equiv 0$ and

$$\frac{dz_j}{dx} = \sum_{p=3}^{n} b_{jp} z_p \qquad j = 3, \cdots, n \tag{4.10}$$

$$\frac{dy_1}{dx} = \sum_{p=1}^{n} c_{1p} \left( z_p + \frac{W_{p2}}{W_{12}} y_1 + \frac{W_{1p}}{W_{12}} y_2 \right) \tag{4.11}$$

$$\frac{dy_2}{dx} = \sum_{p=1}^{n} c_{2p} \left( z_p + \frac{W_{p2}}{W_{12}} y_1 + \frac{W_{1p}}{W_{12}} y_2 \right) \tag{4.12}$$

with

$$b_{jp} = c_{jp} - \frac{W_{j2}}{W_{12}} c_{1p} - \frac{W_{1j}}{W_{12}} c_{2p}.$$

Setting

$$y_1 = p_1 x_{11} + p_2 x_{12}, \qquad y_2 = p_1 x_{21} + p_2 x_{22} \tag{4.13}$$

in (4.10) and (4.11) we find, using

$$dx_{j1}/dx = \textstyle\sum_{p=1}^{n} c_{jp} x_{p1}$$

and

$$dx_{j2}/dy = \textstyle\sum_{p=1}^{n} c_{jp} x_{p2},$$

$$p_1 = \sum_{p=3}^{n} \int_{\omega_1}^{x} \frac{z_p [c_{1p} x_{22} - c_{2p} x_{12}]}{W_{12}} \, dy \tag{4.14}$$

$$p_2 = \sum_{p=3}^{n} \int_{\omega_2}^{x} \frac{z_p [c_{1p} x_{21} - c_{2p} x_{11}]}{W_{12}} \, dx,$$

$\omega_1$ and $\omega_2$ are constants.

From (4.9), (4.10), (4.13) and (4.14) the proof of the lemma for $W_{12}$ everywhere different from zero follows immediately.

Now if $W_{12}$ vanishes at isolated points, the representation (4.6) still holds elsewhere. By (4.9) any solution $z_j$ of (4.10) corresponds to a particular solution of (4.2). Since any solution of (4.2) is single-valued $z_j$ must be single-valued and finite except for poles at $W_{12} = 0$ by (4.9). Furthermore, a solution $x_{pr}$ is defined uniquely by its initial values since (4.2) has no singular points. By (4.5) and (4.6), for $j \geqq 3$

$$x_{jr}(\omega_{1r}) = -x_{j2}(\omega_{1r}) \sum_{p=3}^{n} \int_{\omega_{2r}}^{\omega_{1r}} \frac{z_{pr}[c_{1p}x_{21} - c_{2p}x_{11}]}{W_{12}}\, dx + z_{jr}(\omega_{1r})$$

and a similar equation holds for $x_{jr}(\omega_{2r})$. These initial values are fixed if we fix $z_{jr}(\omega_{1r})$ and the path of integration. Let us fix $z_{jr}(\omega_{1r})$ and denote any two solutions resulting from different paths of integration by $x^1_{jr}$ and $x^2_{jr}$. Then for $x = \omega_{1r}$ and $x = \omega_{2r}$

$$x^1_{jr}(\omega_{1r}) - x^2_{jr}(\omega_{1r}) = x_{j2}(\omega_{1r})M_r$$
$$x^1_{jr}(\omega_{2r}) - x^2_{jr}(\omega_{2r}) = x_{j1}(\omega_{2r})N_r$$

where $M_r$ and $N_r$ are residues from the poles at $W_{12} = 0$. Now $x^1_{jr}(x) - x^2_{jr}(x)$ is a solution of (4.1) satisfying given initial conditions at $\omega_{1r}$ and $\omega_{2r}$. It must in fact be given by

$$x^1_{jr}(x) - x^2_{jr}(x) = x_{j2}(x)M_r = x_{j1}(x)N_r$$

for all $j \geqq 3$, $r$, and $x$ for which $W_{12}(x) \neq 0$. Then for any $r$ either $M_r = N_r = 0$ and $x^1_{jr}(x) = x^2_{jr}(x)$, or $W_{jp} = 0$ for $j \geqq 3, p \geqq 3, x \neq \bar{x}$. By continuity if $W_{jp} = 0$ for $x \neq \bar{x}$, $W_{jp} = 0$ for $x = \bar{x}$. But $W_{jp} = 0, j \geqq 3,\ p \geqq 3$ and $W_{12} = 0$ would imply that the two solutions $x_{j1}$ and $x_{j2}$ are not linearly independent everywhere as assumed. Hence $x^1_{jr}(x) = x^2_{jr}(x)$ for all $r$ and the solution (4.6) is uniquely defined and independent of the path of integration.

We shall need an additional lemma relating the formal solutions of (4.2) with the formal solutions of the reduced system (4.8).

LEMMA 3. The formal equivalent of the reduced system (4.8) found by replacing $x_{j1}$ by $e^{\lambda P_1}\sum_{\nu=0}^{\infty} f_{j1\nu}(x)\lambda^{-\nu}$ and $x_{j2}$ by $e^{\lambda P_2}\sum_{\nu=0}^{\infty} f_{j2\nu}(x)\lambda^{-\nu}$ can be solved formally by series of the form

$$(4.15) \qquad e^{\lambda P_p(x)}\sum_{\nu=0}^{\infty} g_{jp\nu}(x)\lambda^{-\nu} \qquad p = 3, \cdots n, \qquad k = 3, \cdots n,$$

in any domain for which $W_{12}$ does not vanish formally. Furthermore, the expressions (4.15) are related to (4.3) by the formal equivalent of (4.5) and (4.6). The proof follows immediately from Lemma 2 and (4.9).

If $W_{12}$ vanishes for particular values of $x$ in $T$ the formal solutions (4.15) will still exist except at certain values of $x$, where, in fact, each term will have a pole. If $W_{12}$ vanishes identically we reorder the equations since at least one $W_{jp}$ does not vanish identically. For otherwise the series

$$e^{\lambda P_1(x)}\sum_{\nu=0}^{\infty} f_{j1\nu}(x)\lambda^{-\nu} \quad \text{and} \quad e^{\lambda P_2(x)}\sum_{\nu=0}^{\infty} f_{j2\nu}(x)\lambda^{-\nu}$$

would not be formally independent and we would not have $n$ formally independent solutions to start out with.

**5. Subdivisions of $x$-plane in Case I, arg $a'(0) \neq$ arg $b'(0)$.** In order to apply Lemma 1 to (3.1) to find where the formal solutions with exponent $\lambda Q$ are valid we will need to know the behavior of Re $\lambda Q$ and Re $\lambda(Q \pm R)$ in the $x$-plane. Consider first the exponent Re $\lambda(Q + R)$. There are three curves, $C_{k,Q+R}$, $k = 1, 2, 3$, along which Re $\lambda(Q + R) = 0$. Note that these curves are independent of $\lambda$ since arg $\lambda$ is kept constant. They meet at the origin at an angle of $\frac{2}{3}\pi$ and are given near the origin by the lines,

$$\arg x - \frac{\pi}{3} + \tfrac{2}{3} \arg \lambda\left(\sqrt{b'(0)} + \sqrt{a'(0)}\right) = 0, \frac{2\pi}{3}, \frac{4\pi}{3}, \tag{5.1}$$

The curves Re $\lambda(Q + R) =$ constant fill the sectors between $C_{k,Q+R}$. Near the origin they are given by

$$\arg x = \tfrac{2}{3} \cos^{-1}(k_1/|x|^{3/2}) - \tfrac{2}{3} \arg \lambda\left(\sqrt{b'(0)} + \sqrt{a'(0)}\right) + \cdots \tag{5.2}$$

where $k_1$ is a real constant. Similarly there are three curves $C_{k,Q-R}$ on which Re $\lambda(Q - R) = 0$. Near the origin we have

$$\arg x - \tfrac{1}{3}\pi + \tfrac{2}{3} \arg \lambda\left(\sqrt{b'(0)} - \sqrt{a'(0)}\right) = 0, \frac{2\pi}{3}, \frac{4\pi}{3}, \tag{5.3}$$

while the curves Re $\lambda(Q - R) =$ constant are given near the origin by

$$\arg x = \tfrac{2}{3} \cos^{-1}(k_2/|x|^{3/2}) - \tfrac{2}{3} \arg \lambda\left(\sqrt{b'(0)} - \sqrt{a'(0)}\right) \cdots \tag{5.4}$$

where $k_2$ is a real constant. These curves are obtained essentially by a rotation about the origin of the corresponding curves for Re $\lambda(Q + R)$.

We assume that any sector bounded by the adjacent curves Re $\lambda(Q \pm R) = 0$ can be mapped into a quadrant of the $\xi_1$, $\xi_2$ plane where $\xi_1 =$ Re $\lambda_0 (Q - R)$,

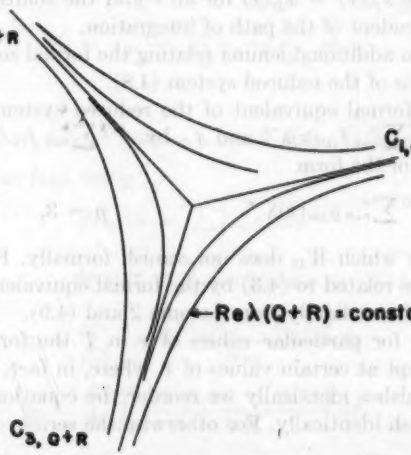



FIG. 1

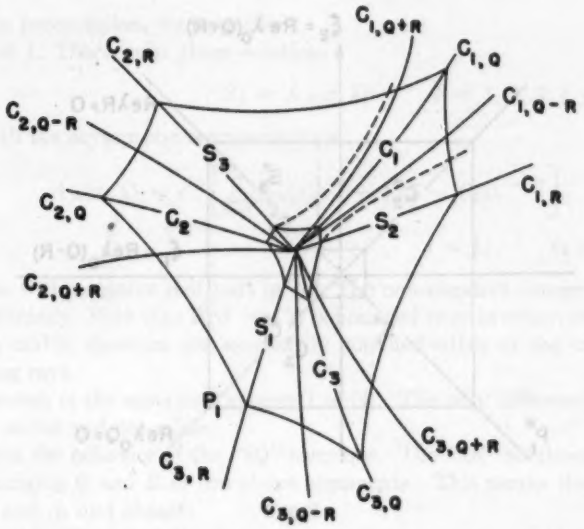



FIG. 2

Dotted line indicates the boundary of $T_\epsilon$.

$\xi_2 = \operatorname{Re} \lambda_0(Q + R)$ where $\lambda_0$ is a fixed value of $\lambda$, $\arg \lambda = \arg \lambda_0$. This is easily seen from (5.2) and (5.3) to be possible near the origin.

Thus the figure cut out by the curves of constant $|\operatorname{Re} \lambda_0(Q \pm R)|$ will be six-sided. Such a figure is chosen to be the boundary of $S$. The constant is determined by the other conditions on $S$, see Section 3. Thus, $S$ is fixed independent of $\lambda$. The corners of $S$ must lie alternately on the curves $\operatorname{Re} \lambda Q = 0$ and $\operatorname{Re} \lambda R = 0$ which we denote by $C_{k,Q}$ and $C_{k,R}$. To fix our ideas we assume that in going counter-clockwise around the origin we run successively, through $C_{1,R}$, $C_{1,Q-R}$, $C_{1,Q}$, $C_{1,Q+R}$, etc.

The lines $C_{k,Q-R}$, $C_{k,Q+R}$ divide $S$ into six open subdomains which we label $C_k$ if the subdomain contains $C_{k,Q}$ and $S_k$ if it contains $C_{k-1,R}$ ($k$ is taken modulo 3). The lines $C_{k-1,R}$ split the $S_k$ domains into two sectors: $S_{k,k+1}$ adjacent to $C_{k+1}$ and $S_{k,k-1}$ adjacent to $C_{k-1}$. Similarly the lines $C_{k,Q}$ split the $C_k$ domains into: $C_{k,k+1}$ adjacent to $S_{k+1}$ and $C_{k,k-1}$ adjacent to $S_{k-1}$.

**6. The "$\lambda Q$" and "$\lambda R$" Solutions.** We now decide which determination of the square root to take for $Q$. We choose $\lambda Q$ to have negative real part in $S_1$. Then $\lambda Q$ has positive real part except in the sector between $C_{2,Q}$ and $C_{3,Q}$. Then $\lambda(Q - R)$ and $\lambda(Q + R)$ also have negative real part in $S_1$. We denote by $P_1$ the corner of $S_1$ that lies on $C_{3,R}$. This will be the point where $|\operatorname{Re} \lambda Q|$ has its maximum.

Now we map $S$ into the $\xi_1$, $\xi_2$-plane where $\xi_1 = \operatorname{Re} \lambda_0(Q - R)$, $\xi_2 = \operatorname{Re} \lambda_0 (Q + R)$. Its image $S^*$ covers a square. In fact, the right half of the square is

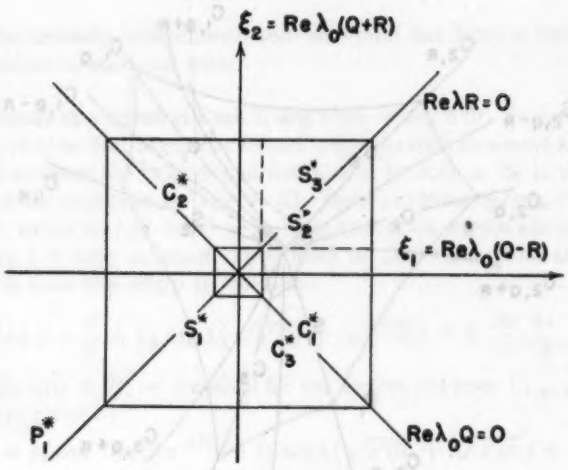



FIG. 3

Dotted line indicates the boundary of $T_\epsilon^*$.

covered twice. Each of $C_k$ and $S_k$ maps into a quadrant. The point $P_1$ maps into $P_1^*$ and $S_1$ maps into the lower left quadrant, $S_1^*$. Going clockwise around the origin from $S_1^*$ we then have $C_2^*$, $S_3^*$ and $C_1^*$. Going counter-clockwise we have successively $C_3^*$ and $S_2^*$.

Now we are able to apply Lemma 1 to (3.1) for the asymptotic expansions containing the exponent $\lambda Q$. In this case $P_1 = -P_2 = \lambda Q = |\lambda/\lambda_0|\lambda_0 Q$, $P_3 = -P_4 = \lambda R$, $P_5 = P_6 = 0$, $T = S$.

From Figure 3 we see that every point in $S_1^* + C_3^* + S_2^*$ can be reached by a path from $P_1^*$, circling below the origin, along which $\xi_1$, $\xi_2$ and $\xi_1 + \xi_2$ are non-decreasing. Every point in $S_1^* + C_2^* + S_3^*$ can be reached by a path above the origin, satisfying the same condition. However, the quadrant $C_1^*$ cannot be reached by such a path. We may expect to apply Lemma 1 then in $S - C_3$. The complete argument is as follows:

Consider the closed subset $T_\epsilon$ of $S - C_3$ which satisfies the conditions

$$\xi_2 > \epsilon > 0 \qquad \text{if } x \subset S_2$$

$$\xi_1^2 + \xi_2^2 > \epsilon^2 \qquad \text{if } x \subset C_2 + S_1 + C_3$$

$$\xi_1 > \epsilon > 0 \qquad \text{if } x \subset S_3.$$

From Figure 3 we see that Lemma 1 can be applied to (3.1) and the domain $T_\epsilon$ for arbitrary $\epsilon > 0$. It follows that there exists a solution $A_{j1}$ of (3.1) with

$$A_{j1} = e^{\lambda Q}\left[\sum_{\nu=0}^{\nu_1} f_{j1\nu}(x)\lambda^{-\nu} + \frac{E(S-\bar{C}_1)}{\lambda^{\nu_1+1}}\right],$$

where $\bar{C}_1$ is the closure of $C_1$.

By cyclic permutation, we have then

THEOREM 1. There exist three solutions

$$Z_j = A_{jk}(x, \lambda) \qquad k = 1, 2, 3; j = 1, \cdots, 6$$

of (3.1) with the asymptotic representations

$$A_{jk}(x, \lambda) = e^{\lambda Q}\left[\sum_{\nu=0}^{\nu_1} f_{j1\nu}(x)\lambda^{-\nu} + E(S - C_k)\lambda^{-\nu_1-1}\right] \tag{6.1}$$

$$j = 1, \cdots, 6; k = 1, 2, 3.$$

$\lambda Q$ is taken with negative real part in $S_k$. The non-negative integer $\nu_1$ may be chosen arbitrarily. Note that $E(S - C_k)$ is bounded only in every closed subset of $S - C_k$ and is therefore not necessarily bounded either at the origin or on its bounding rays.

This theorem is the same as Theorem 1 of [4]. The only difference is that $C_k$ is a whole sector and not a line.

(6.1) gives the behavior of the "$\lambda Q$" solutions. The "$\lambda R$" solutions are found by interchanging $Q$ and $R$ in the above arguments. This means that we interchange $C_k$ and $S_k$ and obtain:

THEOREM 2. There exist solutions

$$Z_1 = B_{jk}(x, \lambda) \qquad k = 1, 2, 3; j = 1, \cdots, 6$$

of (3.1) with the asymptotic representations

$$B_{jk}(x, \lambda) = e^{\lambda R}\left[\sum_{\nu=0}^{\nu_1} f_{j2\nu}(x)\lambda^{-\nu} + E(S - S_k)\lambda^{-\nu_1-1}\right] \tag{6.2}$$

$$j = 1, \cdots, 6; k = 1, 2, 3$$

$\lambda R$ is taken with negative real part in $C_k$.

**7. Complete Set of Solutions.** We note that in any set of three adjacent sectors we will have by Theorems 1 and 2 three asymptotically independent solutions. For example, in every closed subset of $S_1 + C_3 + S_2$ we have $A_{j1}$, $A_{j2}$ and $B_{j3}$. (We include in $S_1 + C_3 + S_2$ the common boundary of $S_1$ and $C_3$ and of $C_3$ and $S_2$.) We now seek three more asymptotic solutions of (3.1) which, along with $A_{j1}$, $A_{j2}$ and $B_{j3}$, will form a complete set in some subdomain of $S_1 + C_3 + S_2$.

To obtain these solutions we apply Lemma 2 to the system (3.1) using the two known solutions $A_{j1}$ for $x_{j1}$ and $A_{j2}$ for $x_{j2}$. Then, by (4.8), the reduced system is

$$\frac{dz_j}{dx} = \sum_{p=3}^{6} \beta_{jp} z_p \qquad j = 3, \cdots, 6 \tag{7.1}$$

with

$$\beta_{jp} = \gamma_{jp} - \frac{W_{j2}}{W_{12}}\gamma_{1p} - \frac{W_{1j}}{W_{12}}\gamma_{2p} \qquad j = 3, \cdots 6; p = 3, \cdots, 6$$

where

$$W_{jp} = \begin{vmatrix} A_{j1} & A_{j2} \\ A_{p1} & A_{p2} \end{vmatrix}.$$

From (6.1) we can determine the asymptotic behavior of $W_{jp}$ in $S_1 + C_3 + S_2$. We exclude from $S$ arbitrarily small but fixed neighborhoods of the points where the leading term of $W_{12}$ vanishes if such points exist.

We wish to apply Lemma 1 to (7.1). By Lemma 3 we have four formal solutions $e^{\lambda P_r}[\sum_{\nu=0}^{\infty} \tau_{jr\nu}(x)\lambda^{-\nu}]$, $r = 3, \cdots, 6$, $j = 3, \cdots, 6$ that satisfy (7.1) with $P_3 = \lambda R$, $P_4 = -\lambda R$, $P_5 = P_6 = 0$. Also $T = S_1 + C_3 + S_2$.

If we choose $\lambda R$ to have negative real part in $C_3$ then, we may use Fig. 3 to see that every point in every closed subset of $S_1 + C_3 + S_2$ satisfies the conditions of Lemma 1 for (7.1) and we have

LEMMA 4. There exists a solution $D_{j3}$ of (7.1) with

$$D_{j3}(x, \lambda) = e^{\lambda R}\left[\sum_{\nu=0}^{\nu_1} \tau_{j3\nu}(x)\lambda^{-\nu} + \frac{E(S_1 + C_3 + S_2)}{\lambda^{\nu_1+1}}\right] \tag{7.2}$$

where $\lambda R$ has negative real part in $C_3$, $W_{12}(x) \neq 0$.

If we use this solution in Lemma 2 to find a solution of the sixth order system (3.1) from the fourth order system (7.1) we will only be led back to $B_{j3}$. However, we do need it for further reductions of order.

Alternatively, $\lambda R$ can be taken with negative real part either in $C_1$ or $C_2$. To fix our ideas, let $\lambda R$ have negative real part in $C_1$ and hence in $S_{21}$. Starting from any point in $S_1 + C_3 + S_2$ we see that if we cross $C_{3R}$, $\operatorname{Re} \lambda R$ must decrease somewhere since $\operatorname{Re} \lambda R = 0$ on $C_{3R}$ and is positive on both sides of it. We expect that the correct domain for $T^*$ will be in this case $S_{12} + C_3 + S_2$. We can prove this either by the argument used in [4] for the case of one exponential exponent or by studying the curves $\lambda R =$ constant in the $x$-plane (by Figure 3, say). We then have

LEMMA 5. There exists a solution $D_{j1}$ of (7.1) with

$$D_{j1}(x, \lambda) = e^{\lambda R}\left[\sum_{\nu=0}^{\nu_1} \tau_{j3\nu}(x)\lambda^{-\nu} + \frac{E(S_{13} + C_3 + S_2)}{\lambda^{\nu_1+1}}\right]. \tag{7.3}$$

where $\lambda R$ has negative real part in $C_1$ and $W_{12}(x) > \delta$ with $\delta$ arbitrarily small.

We repeat this process once again and reduce (7.1) by Lemma 2 using $D_{j1}$ for $x_{j1}$ and $D_{j3}$ for $x_{j2}$. We obtain the system

$$\frac{du_j}{dx} = \sum_{p=5,6} \alpha_{jp} u_p \qquad j = 5, 6$$

$$\alpha_{jp} = \beta_{jp} - \frac{W_{j4}}{W_{34}}\beta_{3p} - \frac{W_{3j}}{W_{34}}\beta_{4p} \tag{7.4}$$

where

$$W_{jp} = \begin{vmatrix} D_{j3} & D_{j1} \\ D_{p3} & D_{p1} \end{vmatrix} \tag{7.5}$$

is known asymptotically in $S_{12} + C_2 + S_2$ through Lemmas 4 and 5. In applying Lemma 1 again we have $P_5 = P_6 = 0$ by Lemma 3, $T = S_{12} + C_2 + S_2$. Thus there exist two solutions $F_{jk}(x, \lambda)$ $j = 5, 6$, $k = 5, 6$ of (7.4) that have asymptotic representations,

$$F_{jk}(x, \lambda) = \sum_{\nu=0}^{\nu_1} \mu_{jk\nu}(x)\lambda^{-\nu} + E(S_{12} + C_2 + S_2)\lambda^{-\nu_1-1} \qquad j = 5, 6; k = 5, 6 \tag{7.6}$$

Now we work backwards with Lemma 2 to reach the sixth order system again. First we determine two more solutions of (7.1) using the solutions $F_{jk}$, (7.6), of (7.4). This yields the "power series" solutions of (7.1), $G_{jk}$, $j = 3, 4, 5, 6$, $k = 5, 6$ with

$$\begin{aligned} G_{3k} &= D_{32}p_{2k} + D_{31}p_{1k} \\ G_{4k} &= D_{42}p_{2k} + D_{41}p_{1k} \qquad k = 5, 6 \qquad j = 5, 6 \\ G_{jk} &= D_{j2}p_{2k} + D_{j1}p_{1k} + F_{jk} \end{aligned} \tag{7.7}$$

where

$$\begin{aligned} p_{2k} &= \sum_{p=5,6} \int_{\omega_1}^{x} \frac{F_{pk}[\beta_{3p}D_{41} - \beta_{4p}D_{31}]}{W_{34}}\,dx \\ p_{1k} &= \sum_{p=5,6} \int_{\omega_2}^{x} \frac{F_{pk}[\beta_{3p}D_{42} - \beta_{4p}D_{32}]}{W_{34}}\,dx; \end{aligned} \tag{7.8}$$

$\omega_1$ is chosen as the point in $S_{21}$ where $\operatorname{Re} \lambda R$ has its minimum when $\operatorname{Re} \lambda R$ is negative in $C_1$. The constant $\omega_2$ is the value of $x$ at the point in $C_2$ where $\operatorname{Re} \lambda R$ has its minimum when $\operatorname{Re} \lambda R$ is negative in $C_3$.

If we allow $x$ to range over the domain $S_{12} + C_2 + S_2$ only, all functions occurring in the integrands of (7.8) have known asymptotic expansions, given in the following way: $F_{pk}$ through (7.6), $D_{pk}$ through (7.2) and (7.3), $W_{34}$ through (7.2) and (7.3) in (7.5), $\beta_{jp}$ through

$$\beta_{jp} = \gamma_{jp} - \frac{W_{j2}}{W_{12}}\gamma_{1p} - \frac{W_{1j}}{W_{12}}\gamma_{2p}.$$

In the Wronskian $W_{jp}$ the asymptotic expansion of (6.1) is used. Again, if $W_{34}$ vanishes we exclude from $S$ fixed neighborhoods of such points. The integrand in the expression for $p_{2k}$ will contain the exponential factor $e^{\lambda R}$ and a power series in $1/\lambda$. Moreover $\lambda R$ is determined by $D_{k1}$ and has negative real part in $C_1$. The end point $x$ of the integral can therefore be reached from $\omega_1$ by a path along which $\operatorname{Re} \lambda R$ is non-decreasing. We may therefore integrate by parts to find the asymptotic expansion of $p_{2k}$ using the fact that the remainder terms are bounded in $\delta$ and of lower order in $\lambda$. A similar treatment gives the asymptotic expansion of $p_{1k}$.

We find after some calculations an expression for $G_{jk}$ of the form

$$G_{jk} = \sum_{\nu=0}^{\nu_1} \tau_{jk\nu}(x)\lambda^{-\nu} + E(S_{12} + C_2 + S_2)\lambda^{-\nu_1-1} \qquad j = 3, 4, 5, 6; k = 5, 6. \tag{7.9}$$

Now we apply Lemma 2 again to the system (7.1) using the solutions $G_{jk}$ of (7.9), along with the known solutions $A_{j1}$ and $A_{j2}$. This yields two "power-series" solutions of (3.1).

We omit the rather lengthy details since the method is exactly the same as we used to obtain (7.9). We find then two solutions

$$U_{jk} = \sum_{\nu=0}^{\nu_1} f_{jk\nu}(x)\lambda^{-\nu} + \frac{E(S_{12} + C_3 + S_2)}{\lambda^{\nu_1+1}} \qquad j = 1, 2, \cdots, 6 \qquad k = 5, 6,$$

valid say for $|W_{12}(x)| > \delta$, $|\overline{W}_{34}(x)| > \delta$ where $\delta$ is arbitrarily small but independent of $\lambda$.

One more solution is still to be found using the solution $D_{j1}$ in Lemma 2. We omit the details and mention only the pertinent facts. The exponents which occur in the integrals corresponding to (4.6) turn out to be $e^{\lambda(R+Q)}$ and $e^{\lambda(R-Q)}$ where $\lambda R$ has negative real part in $C_1$ and hence in $S_{21}$ while $\lambda Q$ has negative real part in $S_2$. This means that $\lambda(R + Q)$ has negative real part in $S_2$ and that $\lambda(R - Q)$ has negative real part in $S_{13}$. Hence we choose as constants of integration for (4.6) the points in $S_2$ and $S_{13}$ where $\operatorname{Re}\lambda(R + Q)$ and $\operatorname{Re}\lambda(R - Q)$ respectively have minima. Then the real parts of the exponents in the integrands are non-decreasing and we may integrate to obtain the asymptotic expansions. This yields the solution

$$B_j^1 = e^{\lambda R}\left[\sum_{\nu=0}^{\nu_1} f_{j1\nu}(x)\lambda^{-\nu} + \frac{E(S_{13} + C_3 + S_2)}{\lambda^{\nu_1+1}}\right],$$

where $\lambda R$ has negative real part in $C_1$.

*We now have a complete set of six solutions whose asymptotic behavior is given in $S_{13} + C_3 + S_2$ by the formal solutions* (3.3)*of the sixth order system* (3.1).

Now we might also have chosen $\lambda R$ with negative real part in $C_2$. Then we have a solution

$$B_j^2 = e^{\lambda R}\sum_{\nu=0}^{\nu_1} f_{j1\nu}(x)\lambda^{-\nu} + \frac{E(S_1 + C_3 + S_{23})}{\lambda^{\nu_1+1}}\Bigg]$$

where $\lambda R$ has negative real part in $C_2$. Similar solutions of the power series type $U_{jk}$ can also be found which are valid in $S_1 + C_3 + S_{23}$.

By cyclic permutation we find

THEOREM 3. A complete set of six independent solutions of (3.1) is represented by (3.3) in any of the following domains:

(a) In any closed subset of $S_{k-1,k} + C_k + S_{k+1}$ $\qquad k = 1, 2, 3.$

For the first and second $\lambda Q$-solutions, $\lambda Q$ is chosen to have negative real part in $S_{k+1}$ and $S_{k-1}$ respectively. The first solution can be continued around in either direction to $C_{k+1}$, the second to $C_{k-1}$. For the two $\lambda R$-solutions, $\lambda R$ has negative real part in $C_k$ and $C_{k+1}$ respectively. The first solution can be continued around in either direction to $S_k$.

(b) In any closed subset of $S_{k-1} + C_k + S_{k+1,k}$ $\qquad k = 1, 2, 3.$

For the two $\lambda Q$-solutions see (a). For the two $\lambda R$-solutions $\lambda R$ is determined by the condition that it has negative real part in $C_k$ and $C_{k-1}$ respectively. The first of these can be continued around in either direction to $S_k$.

In a strictly analogous way we may find other domains of validity for a complete set of solutions. At the beginning of Section 7 we chose the domain $S_1 + C_2 + S_2$ and the basic $\lambda Q$-solutions $A_{f1}$ and $A_{f2}$ whose asymptotic expansions are valid in $S_1 + C_2 + S_2$. We could have chosen alternatively $C_1 + S_3 + C_3$ with the basic $\lambda R$ solutions $B_{f1}$ and $B_{f2}$. By interchanging the roles of $Q$ and $R$ we find

THEOREM 4. A complete set of six independent solutions of (3.1) is represented by (3.3)

(c) In any closed subset of $C_{k-1,k} + S_k + C_{k+1}$, $k = 1, 2, 3.$

For the first and second $\lambda R$-solutions, $\lambda R$ is chosen to have negative real part in $C_{k+1}$ and $C_{k-1}$ respectively. The first solution can be continued around in either direction to $S_{k+1}$, the second to $S_{k-1}$. For the two $\lambda Q$-solutions, $\lambda Q$ has negative real part in $S_k$ and $S_{k+1}$ respectively. The first solution can be continued in either direction to $C_k$.

(d) In any closed subset of $C_{k-1} + S_k + C_{k+1,k}$, $k = 1, 2, 3.$

For the two $\lambda R$-solutions see (c). For the two $\lambda Q$-solutions $\lambda Q$ is determined by the condition that it has negative real part in $S_k$ and $S_{k-1}$ respectively. The first of these solutions can be continued around in either direction to $C_k$.

Combining Theorems 3 and 4 we obtain finally:

THEOREM 5. Corresponding to any five adjacent sectors selected from the twelve sectors $C_{kp}$ and $S_{kp}$ there exists a complete set of six solutions of (3.1) whose asymptotic behavior is given by (3.3) in the five open sectors.

In solving any boundary value problem we will have to decide which set of five adjacent sectors include the end points given by the problem.

In the problem of stability $C_{k,R}$, $C_{k,Q-R}$, $C_{k,Q}$ and $C_{k,Q+R}$ lie very closely together for each $k$. Theorem (3a) gives the most suitable domain for many purposes. It stretches around the origin through almost 240°.

It should be noted that the functions $E(T)$ are bounded only in every closed subdomain of $T$. Hence the regions of validity for the asymptotic expressions always exclude an arbitrary but finite neighborhood of the origin and an arbitrary neighborhood of the boundaries formed by the outside sector edges. The present methods cannot lead to any results about the asymptotic behavior at the origin or in a neighborhood of the origin which shrinks to zero as $\lambda \to \infty$.

This problem has been studied for the fourth order equation by Wasow [8].

**8. Case II** $\arg a'(0) = \arg b'(0)$. It does not seem possible to apply the present methods of finding the domains of validity to $\arg a'(0) = \arg b'(0)$ unless $\arg a'(x) = \arg b'(x)$. In general the lines $\operatorname{Re} \lambda(Q \pm R) =$ constant intersect each other within each sector more than once and it is not clear that one can find the necessary paths along which $\operatorname{Re} \lambda(Q + R)$ and $\operatorname{Re} \lambda(Q - R)$ are monotonic non-decreasing.

For the case in which $\arg a'(x) \equiv \arg b'(x)$ we have $R = \alpha Q$ where $\alpha$ is a real positive constant which we may take less than 1. This is, for many practical purposes, the case in the stability problem for the Prandtl number $\sigma_0$ is almost a constant, about .75, cf. (2.16) and (3.4). We will have then $\operatorname{Re} \lambda R = \alpha \operatorname{Re} \lambda Q$ and all the lines $C_{k,Q}$, $C_{k,R}$ etc. coincide. This case can be handled by a straight forward application of Wasow's methods [4]. The results are given by Theorems 1, 2 and 3 (a) if we consider that the domains $C_k$ have degenerated into lines. In particular in analogy to Theorem 3 we have:

THEOREM 6. If $R = \alpha Q$ where $\alpha$ is a real positive constant less than 1 there exist three complete sets of solutions of (3.1) whose asymptotic behavior is given by (3.3) in each of the domains $S - S_k$. For one $\lambda Q$-solution and one $\lambda R$-solution, $\lambda Q$ and $\lambda R$ have negative real part in $S_{k-1}$. For the other $\lambda Q$ and $\lambda R$ solutions, $\lambda Q$ and $\lambda R$ have negative real part in $S_{k+1}$.

**9. The Continuation of the "Power-Series" Solutions.** Theorems 1 to 6 tell us where the asymptotic forms (3.3) can be used to represent solutions. They do not tell us what is the largest domain of validity for the asymptotic expansion of any particular solution. This problem can not be settled by the present methods. Some useful results can be obtained beyond those given in Theorems 3 and 4. We can extend the solutions which behave like power series in any given five sectors to one additional sector. Furthermore, if the leading terms of the power series are not single valued in $x$ then the corresponding solution diverges in at least four sectors. This last result is of particular importance and establishes the existence of an inner viscous layer since equations (2.14) are singular at $w = c$ and in fact have one regular and one singular solution.

These results are expressed in two theorems:

THEOREM 7: There exists a solution of (3.1) which can be represented asymptotically by power series in $1/\lambda$ in a set of six adjacent sectors selected from the twelve sectors $S_{k_2}$ and $C_{k_2}$ provided the six sectors together are bounded on one side by one of the curves $C_{k,Q}$ and on the other side by one of $C_{k,R}$.

THEOREM 8. Consider a solution of (3.1) which is represented asymptotically by one of the power series (3.3), $j = 5, 6$, in six adjoining sectors. If the leading coefficient of this series is multi-valued in a neighborhood of the origin then the given solution does not converge to a finite value as $\lambda \to \infty$ at any point of at least four of the remaining sectors. These four sectors lie next to the given six.

We introduce a new notation $P_k$, $k = 1, 2, \cdots, 12$ for the twelve sectors in the $x$-plane and the subscript $k$ to denote the twelve sets of solutions which correspond to each group of five adjacent sectors. To fix the ideas let $P_1$ lie between $C_{1,Q+R}$ and $C_{2,R}$, see Figure 4. We denote with the second subscript $k$ a complete set of six independent solutions whose asymptotic behavior is given by (3.3) in the five adjacent sectors $P_{k-2} + P_{k-1} + P_k + P_{k+1} + P_{k+2}$. $A_{jk}$ and $\tilde{A}_{jk}$ will denote the components of the two $\lambda Q$-solutions, $B_{jk}$ and $\tilde{B}_{jk}$ the two $\lambda R$-solutions and $U_{jk}$ and $\tilde{U}_{jk}$ the two "power-series" solutions. Here $j = 1, 2, \cdots, 6$, $k = 1, 2, \cdots, 12$.

We shall first prove Theorem 7 for $U_{j7}$. That is we have

$$U_{j7} = \sum_{\nu=0}^{\nu_1} f_{j\nu} \lambda^{-\nu} + \lambda^{-\nu_1-1} E(P_5 + P_6 + P_7 + P_8 + P_9) \tag{9.1}$$

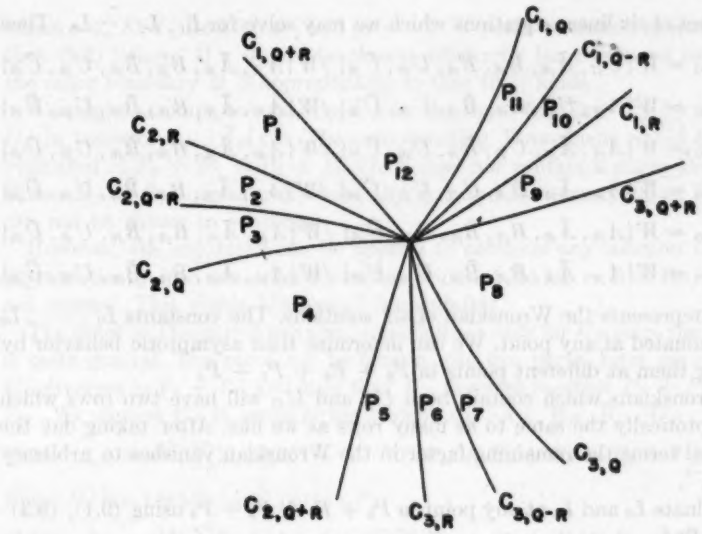



FIG. 4

and we shall show that

$$U_{j7} = \sum_{\nu=0}^{r_1} f_{j3\nu}\lambda^{-\nu} + \lambda^{-r_1-1}E(P_4 + P_5 + P_6 + P_7 + P_8 + P_9). \tag{9.2}$$

We could also use $f_{j4\nu}$ instead of $f_{j3\nu}$.

We express $U_{j7}$ as a linear combination of the six solutions whose asymptotic expansions are known in $P_4 + P_5 + P_6 + P_7 + P_8$, i.e., the solutions $A_{j6}$, $\tilde{A}_{j6}$, $B_{j6}$, $\tilde{B}_{j6}$, $U_{j6}$, and $\tilde{U}_{j6}$. We have from (3.3) and Theorem 5:

$$A_{j6} = e^{\lambda Q}\{\sum_{\nu=0}^{r_1} f_{j1\nu}(x)\lambda^{-\nu} + \lambda^{-r_1-1}E(\sum_4^8 P_k)\} \tag{9.3}$$

where $\lambda Q$ has negative real part in $P_8$,

$$\tilde{A}_{j6} = e^{\lambda Q}\{\sum_{\nu=0}^{r_1} f_{j2\nu}(x)\lambda^{-\nu} + \lambda^{-r_1-1}E(\sum_4^8 P_k)\} \tag{9.4}$$

where $\lambda Q$ has negative real part in $P_4 + P_5 + P_6 + P_7$,

$$B_{j6} = e^{\lambda R}\{\sum_{\nu=0}^{r_1} f_{j3\nu}(x)\lambda^{-\nu} + \lambda^{-r_1-1}E(\sum_4^8 P_k)\} \tag{9.5}$$

where $\lambda R$ has negative real part in $P_6 + P_7 + P_8$,

$$\tilde{B}_{j6} = e^{\lambda R}\{\sum_{\nu=0}^{r_1} f_{j4\nu}(x)\lambda^{-\nu} + \lambda^{-r_1-1}E(\sum_4^8 P_k)\} \tag{9.6}$$

where $\lambda R$ has negative real part in $P_4 + P_5$,

$$U_{j6} = \sum_{\nu=0}^{r_1} f_{j5\nu}(x)\lambda^{-\nu} + \lambda^{-r_1-1}E(\sum_4^8 P_k) \tag{9.7}$$

$$\tilde{U}_{j6} = \sum_{\nu=0}^{r_1} f_{j6\nu}(x)\lambda^{-\nu} + \lambda^{-r_1-1}E(\sum_4^8 P_k) \tag{9.8}$$

where $j = 1, 2, \cdots, 6$. Then we set

$$U_{j7} = L_1A_{j6} + L_2\tilde{A}_{j6} + L_3B_{j6} + L_4\tilde{B}_{j6} + L_5U_{j6} + L_6\tilde{U}_{j6} \tag{9.9}$$

$$j = 1, \cdots, 6.$$

(9.9) is a set of six linear equations which we may solve for $L_1, L_2 \cdots L_6$. Then

$$
\begin{aligned}
L_1 &= W\{U_{j7}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} / W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} \\
L_2 &= W\{A_{j8}, U_{j7}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} / W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} \\
L_3 &= W\{A_{j8}, \tilde{A}_{j8}, U_{j7}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} / W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} \\
L_4 &= W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, U_{j7}, U_{j8}, \tilde{U}_{j8}\} / W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} \\
L_5 &= W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j7}, \tilde{U}_{j8}\} / W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} \\
L_6 &= W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, U_{j7}\} / W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\}
\end{aligned} \tag{9.10}
$$

where $W$ represents the Wronskian of six solutions. The constants $L_1, \cdots, L_6$ can be evaluated at any point. We can determine their asymptotic behavior by evaluating them at different points in $P_5 + P_6 + P_7 + P_8$.

The Wronskians which contain both $U_{j8}$ and $U_{j7}$ will have two rows which are asymptotically the same to as many rows as we like. After taking out the exponential terms the remaining factor in the Wronskian vanishes to arbitrary order in $\lambda^{-1}$.

We evaluate $L_5$ and $L_6$ at any point in $P_5 + P_6 + P_7 + P_8$ using (9.1), (9.3)–(9.8). We find

$$L_5 = 1 + \lambda^{\nu_2} E, \qquad L_6 = \lambda^{-\nu_2} E \tag{9.11}$$

where $\nu_2$ is an arbitrary positive integer, $E$ is a bounded function of $\lambda$. $L$ is evaluated at the point in $P_5 + P_6 + P_7$ on $C_{2R}$ where $|\operatorname{Re} \lambda Q|$ is a maximum, say $|\lambda| a$. We find at this point

$$
\begin{aligned}
W\{U_{j7}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} &= e^{-|\lambda| a} \lambda^{-\nu_1} E \\
W\{A_{j8}, \tilde{A}_{j8}, B_{j8}, \tilde{B}_{j8}, U_{j8}, \tilde{U}_{j8}\} &= \lambda^{-t} \Delta
\end{aligned}
$$

where $\Delta \neq 0$ since the solutions are asymptotically independent and $t$ is some fixed integer. Hence,

$$L_1 = e^{-|\lambda| a} \lambda^{-\nu_2} E, \tag{9.12}$$

$a > 0$ is the value of $|\operatorname{Re} \lambda(Q \pm R)|$ along the boundary of $S$. $L_2$ is found at any point in $P_8$, and $L_3$ in $P_8$:

$$
\begin{aligned}
L_2 &= e^{-|\lambda| b_2} \lambda^{-\nu_2} E \qquad && b_2 > 0 \\
L_3 &= e^{-|\lambda| b_3} \lambda^{-\nu_2} E \qquad && b_3 > 0.
\end{aligned} \tag{9.13}
$$

$L_4$ is found at the point in $P_5$ where $|\operatorname{Re} \lambda R| = |\lambda| (a/2 - \epsilon)$ say where $\epsilon$ is arbitrarily small. Such a point exists since on $C_{2,Q+R}$ we have $\operatorname{Re} \lambda(Q + R) = 0$ and as we approach $C_{2,Q+R}$ through $P_5$ on the outer boundary of $S$ we have $|\operatorname{Re} \lambda(Q - R)| = |\lambda| a$.

$$L_4 = e^{-|\lambda| (\frac{1}{2} a - \epsilon)} \lambda^{-\nu_2} E \tag{9.14}$$

Substituting (9.11)–(9.14) in (9.9) and using the expansions (9.3)–(9.8) we can evaluate $U_7$ in $P_4$. Since $|\operatorname{Re} \lambda Q| \leqq a$ throughout $S$ we find

$$U_{j7} = \textstyle\sum_{\nu=0}^{\nu_1} f_{j8\nu}(x) \lambda^{-\nu} + \lambda^{-\nu_1 - 1} E(P_4 \cdot S^*) \tag{9.15}$$

where $x \subset S^*$ if $|\operatorname{Re} \lambda R| \leqq |\lambda|\,|\tfrac{1}{2}a - \epsilon|$. If $a$ can be chosen arbitrarily large then (9.2) follows. If $a$ cannot be chosen arbitrarily large then we must shrink the outer boundary of $S$ appropriately so that (9.2) holds.

We might attempt to continue $U_{j7}$ in the same way into $P_{10}$ by expressing $U_{j7}$ in terms of $A_{j8}$, $\tilde{A}_{j8}$ etc. The corresponding Wronskians would have to be evaluated in $P_6 + P_7 + P_8 + P_9$ which does not contain a curve $\operatorname{Re} \lambda R = 0$. Either $B_{j8}$ or $\tilde{B}_{j8}$ increases exponentially in $\lambda$ and the corresponding Wronskian can not be shown to converge.

However, the argument can be applied to continue any solution $U_{j8}$ or $U_{j8}$ asymptotically into one adjacent sector provided $\operatorname{Re} \lambda Q = 0$ or $\operatorname{Re} \lambda R = 0$ is not crossed. This yields Theorem 7 immediately.

The proof of Theorem 8 will also be carried out only for $U_{j7}$ since the argument is quite general. The proof can be broken into two parts. First we prove that $U_{j7}$ diverges in $P_1 + P_{12}$ and later that $U_{j7}$ diverges in either $P_2 + P_3$ or $P_{10} + P_{11}$. We express $U_{j7}$ in terms of the system $A_{j2}$, $\tilde{A}_{j2}$, $B_{j2}$, $\tilde{B}_{j2}$, $U_{j2}$, $\bar{U}_{j2}$ where

$$(9.16)\quad A_{j2} = e^{\lambda Q}\{\textstyle\sum_{\nu=0}^{\nu_1} f_{j1}\lambda^{-\nu} + \lambda^{-\nu_1-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)\}$$

where $\lambda Q$ has negative real part in $P_4$,

$$(9.17)\quad \tilde{A}_{j2} = e^{\lambda Q}\{\textstyle\sum_{\nu=0}^{\nu_1} f_{j2}\lambda^{-\nu} + \lambda^{-\nu_1-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)\}$$

where $\lambda Q$ has negative real part in $P_{12} + P_1 + P_2 + P_3$,

$$(9.18)\quad B_{j2} = e^{\lambda R}\{\textstyle\sum_{\nu=0}^{\nu_1} f_{j3}\lambda^{-\nu} + \lambda^{-\nu_1-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)\}$$

where $\lambda R$ has negative real part in $P_2 + P_3 + P_4$,

$$(9.19)\quad \tilde{B}_{j2} = e^{\lambda R}\{\textstyle\sum_{\nu=0}^{\nu_1} f_{j4}\lambda^{-\nu} + \lambda^{-\nu_1-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)\}$$

where $\lambda R$ has negative real part in $P_1 + P_{12}$,

$$(9.20)\quad U_{j2} = \textstyle\sum_{\nu=0}^{\nu_1} f_{j5}\lambda^{-\nu} + \lambda^{-\nu_1-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)$$

$$(9.21)\quad \bar{U}_{j2} = \textstyle\sum_{\nu=0}^{\nu_2} f_{j6}\lambda^{-\nu} + \lambda^{-\nu_1-1}E(P_{12} + P_1 + P_2 + P_3 + P_4).$$

We set

$$(9.22)\quad U_{j7} = M_1A_{j2} + M_2\tilde{A}_{j2} + M_3B_{j2} + M_4\tilde{B}_{j2} + M_5U_{j2} + M_6\bar{U}_{j2}$$

and evaluate $M_2$, $M_4$, $M_5$ and $M_6$ from the appropriate Wronskians evaluated asymptotically by (9.16)–(9.21) at any point in $P_4$.

We find

$$(9.23)\quad \begin{aligned} M_2 &= e^{-\lambda a_2}\lambda^{-\nu_1}E, \qquad M_4 = e^{-\lambda a_1}\lambda^{-\nu_1}E, \\ M_5 &= 1 + \lambda^{-\nu_1}E, \qquad M_6 = \lambda^{-\nu_1}E \end{aligned}$$

where $a_1 > 0$, $a_2 > 0$ and $\nu_1$ is an arbitrary positive integer. If we try to evaluate $M_1$ and $M_3$ in this way we get positive exponential terms. It follows from (9.22), (9.23) and (9.16)–(9.21) that if $U_{j7}$ converges at any point in $P_1 + P_{12}$, $M_1A_{j2} + M_3B_{j2}$ converges at that point.

Now we suppose that as $\lambda \to \infty$, $U_{j7}$ converges in $\sum_1^9 P_k$ to a multivalued function of $x$ and that $U_{j7}$ converges at a point $x_1$ in $P_1 + P_{12}$. Then $M_1A_{j2} + M_3B_{j2}$ has a limit w.r.t. $\lambda$ at $x_1$ for $j = 1, 2, \cdots, 6$.

Consider the determinants

$$\begin{vmatrix} A_{j2} & B_{j2} \\ A_{p2} & B_{p2} \end{vmatrix}.$$

At least one of them has an asymptotic expansion in $P_1 + P_{12}$ which does not vanish identically and, we may assume, does not vanish for $x = x_1$. We may suppose that for this determinant $j = 1$, $p = 2$. We set:

$$A_{12} = e^{\lambda Q(x)}\lambda^{-\gamma_1}F_1(x, \lambda) \tag{9.24}$$

where

$$F_1(x, \lambda) = \sum_{r=0}^{\mu_1} F_{1r}\lambda^{-r} + \lambda^{-\mu_1-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)$$

$$F_{10}(x) \not\equiv 0 \text{ in } S,$$

$$A_{22} = e^{\lambda Q(x)}\lambda^{-\gamma_2}F_2(x, \lambda) \tag{9.25}$$

where

$$F_2(x, \lambda) = \sum_{r=0}^{\mu_2} F_{2r}\lambda^{-r} + \lambda^{-\mu_2-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)$$

$$F_{20}(x) \not\equiv 0 \text{ in } S,$$

$$B_{12} = e^{\lambda R(x)}\lambda^{-\gamma_3}F_3(x, \lambda) \tag{9.26}$$

where

$$F_3(x, \lambda) = \sum_{r=0}^{\mu_3} F_{3r}\lambda^{-r} + \lambda^{-\mu_3-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)$$

$$F_{30}(x) \not\equiv 0 \text{ in } S,$$

$$B_{22} = e^{\lambda R(x)}\lambda^{-\gamma_4}F_4(x, \lambda) \tag{9.27}$$

where

$$F_4(x, \lambda) = \sum_{r=0}^{\mu_4} F_{4r}\lambda^{-r} + \lambda^{-\mu_4-1}E(P_{12} + P_1 + P_2 + P_3 + P_4)$$

$$F_{40}(x) \not\equiv 0 \text{ in } S.$$

In addition we set

$$M_1A_{12}(x_1) + M_3B_{12}(x_1) = C_1(\lambda), \qquad M_1A_{22}(x_1) + M_3B_{22}(x_1) = C_2(\lambda) \tag{9.28}$$

where $C_1(\lambda)$ and $C_2(\lambda)$ have limits as $\lambda \to \infty$.

If we substitute (9.24)–(9.27) in (9.28) and solve for $M_1e^{\lambda Q(x_1)}$ and $M_3e^{\lambda R(x_1)}$ we obtain

$$M_1 e^{\lambda Q(x_1)} = \begin{vmatrix} C_1 & \lambda^{-\gamma_3}F_3(x_1) \\ C_2 & \lambda^{-\gamma_4}F_4(x_1) \end{vmatrix} \Bigg/ \begin{vmatrix} \lambda^{-\gamma_1}F_1(x_1) & \lambda^{-\gamma_3}F_3(x_1) \\ \lambda^{-\gamma_2}F_2(x_1) & \lambda^{-\gamma_4}F_4(x_1) \end{vmatrix}$$

$$M_3 e^{\lambda R(x_1)} = \begin{vmatrix} \lambda^{-\gamma_1}F_1(x_1) & C_1 \\ \lambda^{-\gamma_2}F_2(x_1) & C_2 \end{vmatrix} \Bigg/ \begin{vmatrix} \lambda^{-\gamma_1}F_1(x_1) & \lambda^{-\gamma_3}F_3(x_1) \\ \lambda^{-\gamma_2}F_2(x_1) & \lambda^{-\gamma_4}F_4(x_1) \end{vmatrix} \tag{9.29}$$

Now

$$\begin{vmatrix} \lambda^{-\gamma_1}F_1(x_1) & \lambda^{-\gamma_3}F_3(x_1) \\ \lambda^{-\gamma_2}F_2(x_1) & \lambda^{-\gamma_4}F_4(x_1) \end{vmatrix}$$

may vanish like some power of $\lambda$ but since the asymptotic expansion of the determinant

$$\begin{vmatrix} A_{11} & B_{11} \\ A_{12} & B_{12} \end{vmatrix}$$

does not vanish identically there is a positive integer $n$ such that

$$\lim_{|\lambda|\to\infty} \lambda^n \begin{vmatrix} \lambda^{-\gamma_1}F_1 & \lambda^{-\gamma_3}F_3 \\ \lambda^{-\gamma_2}F_2 & \lambda^{-\gamma_4}F_4 \end{vmatrix} \neq 0. \tag{9.30}$$

From (9.24)–(9.28) and (9.30) substituted in (9.29) it follows that $M_1 e^{\lambda Q(x_1)}\lambda^{-n}$ and $M_3 e^{\lambda R(x_1)}\lambda^{-n}$ have limits as $|\lambda| \to \infty$.

Now consider any point $x$ in $P_{12} + P_1 + P_2 + P_3 + P_4$ for which

$$\operatorname{Re} \lambda R(x) < \operatorname{Re} \lambda R(x_1) \quad \text{and} \quad \operatorname{Re} \lambda Q(x) < \operatorname{Re} \lambda Q(x_1).$$

For this point

$$\begin{aligned} M_1 A_{j2} + M_3 B_{j2} = {} & (M_1 e^{\lambda Q(x_1)}\lambda^{-n})(e^{\lambda Q(x)-\lambda Q(x_1)}\lambda^n)(e^{-\lambda Q(x)}A_{j2}) \\ & + (M_3 e^{\lambda R(x_1)}\lambda^{-n})(e^{\lambda R(x)-\lambda R(x_1)}\lambda^n)(e^{-\lambda R(x)}B_{j2}). \end{aligned} \tag{9.31}$$

Now $e^{\lambda Q(x)-\lambda Q(x_1)}\lambda^n$ and $e^{\lambda R(x)-\lambda R(x_1)}\lambda^n$ converge uniformly to zero as $|\lambda| \to \infty$ and $e^{-\lambda Q(x)}A_{j2}$ and $e^{-\lambda R(x)}B_{j2}$ converge uniformly to limits as $|\lambda| \to \infty$ by (9.16) and (9.18). Since $M_1 e^{\lambda Q(x_1)}\lambda^{-n}$ and $M_3 e^{\lambda R(x_1)}\lambda^{-n}$ have limits it follows that $M_1 A_{j2} + M_3 B_{j2}$ converges uniformly to zero in any closed subdomain of $P_{12} + P_1 + P_2 + P_3 + P_4$ for which $\operatorname{Re} \lambda R(x) < \operatorname{Re} \lambda R(x_1)$ and $\operatorname{Re} \lambda Q(x) < \operatorname{Re} \lambda Q(x_1)$. Hence $U_{j7}$ converges uniformly to a limit in this subdomain.

Similarly by using the fundamental set of solutions for $P_9 + P_{10} + P_{11} + P_{12} + P_1$ we find that $U_{j7}$ converges uniformly to a limit in any closed subdomain of $P_9 + P_{10} + P_{11} + P_{12} + P_1$ for which $\operatorname{Re} \lambda R(x) < \operatorname{Re} \lambda R(x_1)$ and $\operatorname{Re} \lambda Q(x) < \operatorname{Re} \lambda Q(x_1)$.

From Theorem 7 we know that $U_7$ converges uniformly to a limit in $\sum_1^9 P_k$. Thus $U_{j7}$ converges uniformly in a doubly-connected domain about the origin bounded by $\operatorname{Re} \lambda Q(x) = \operatorname{Re} \lambda Q(x_1)$ and $\operatorname{Re} \lambda R(x) = \operatorname{Re} \lambda R(x_1)$. Thus $U_{j7}$ must converge for each $j = 1, 2, 3, \cdots, 6$ to a single valued function of $x$.

Hence if $U_{j7}$ converges to a multivalued function of $x$, our hypothesis that $U_7$ converges at a point in $P_1 + P_{12}$ is false and thus $U_{j7}$ diverges everywhere in $P_1 + P_{12}$. This result is strictly analogous to that of Wasow [4].

Now we shall prove that $U_{j7}$ diverges everywhere in $P_2 + P_3$ or in $P_{10} + P_{11}$. Suppose $U_{j7}$ converges at a point in $P_2 + P_3$. We express $U_{j7}$ in terms of the fundamental set of $P_2 + P_3 + P_4 + P_5 + P_6$, namely, $A_{j4}$, $\bar{A}_{j4}$, $B_{j4}$, $\bar{B}_{j4}$, $U_{j4}$

and $\bar{U}_{jk}$. We find after some computation that $U_{j7}$ converges in $P_2 + P_3$ if, and only if, $MA_{jk}$ converges in $P_2 + P_3$, where

$$M = W\{U_{j7}, \bar{A}_{jk}, B_{jk}, \bar{B}_{jk}, U_{jk}, \bar{U}_{jk}\}/ \tag{9.32}$$
$$W\{A_{jk}, \bar{A}_{jk}, B_{jk}, \bar{B}_{jk}, U_{jk}, \bar{U}_{jk}\}$$

$$A_{jk} = e^{\lambda Q}\{\textstyle\sum_{\nu=0}^{\nu_1} f_{j1\nu}\lambda^{-\nu_1-1}E(\sum_2^6 P_k)\} \tag{9.33}$$

where $\lambda Q$ has negative real part in $P_4 + P_5 + P_6$ and not all of $f_{j1\nu}$ vanish identically. Suppose $U_{j7}$ converges at $x_2 \subset P_2 + P_3$. Then from (9.32) we see that $Me^{\lambda Q(x_2)}\lambda^{-n}$ has a limit as $\lambda \to \infty$, where $n$ is some fixed integer. It follows that $MA_{jk}$ converges uniformly to zero in any closed subdomain for which $\operatorname{Re} \lambda Q(x) < \operatorname{Re} \lambda Q(x_2)$. It is not difficult to prove then that

$$U_{j7} = \textstyle\sum_{\nu=0}^{\nu_1} f_{j1\nu}\lambda^{-\nu} + \lambda^{-\nu_1-1}E(\overline{P_2 + P_3}) \text{ where } \overline{P_2 + P_3} \tag{9.34}$$

is the subdomain of $P_2 + P_3$ for which $\operatorname{Re} \lambda Q(x) < \operatorname{Re} \lambda Q(x_2)$. We may now apply the reasoning of the first part of Theorem 8 to $U_{j7}$ in the domain $\overline{P_2 + P_3} + P_4 + P_5 + P_6 + P_7$ to show that $U_{j7}$ diverges everywhere in $P_{10} + P_{11}$ since $U_{j7}$ does not converge to a set of single-valued functions.

Similarly if $U_{j7}$ converges at a point in $P_{10} + P_{11}$, $U_{j7}$ diverges in $P_2 + P_3$. We have finally proved that $U_{j7}$ diverges in $P_{12} + P_1 + P_2 + P_3$ or in $P_{10} + P_{11} + P_{12} + P_1$. It may of course diverge in both.

The same reasoning applies to $U_{jk}$ for $k = 1, 2, \cdots, 12$. This yields Theorem 8.

**10. The Stability Problem.** The results obtained in Theorems 1–8 will now be applied to the stability problem of Section 2.

First of all, comparing (2.16) and (3.4) we see that the origin of the general system corresponds to the point $y_0$ where $w(y_0) = c$ and we may set $Q = Q_1$, $R = Q_2$, i.e.,

$$b(x) = \frac{i}{\nu_1(y)}\,(w(y) - c), \qquad a(x) = \frac{i\sigma_0(y)}{\nu_1(y)}\,(w(y) - c).$$

The twelve curves which limit the stability of the asymptotic expressions are accordingly, $\operatorname{Re} \lambda Q_1 = 0$, $\operatorname{Re} \lambda Q_2 = 0$, $\operatorname{Re} \lambda(Q_1 \pm Q_2) = 0$ as might be expected from the singularities involved in (2.16). They are given approximately by

$$\pi/4 + \tfrac{1}{2} \arg w'(y_0) - \tfrac{1}{2} \arg \nu_1(y_0) + \arg \lambda + \tfrac{3}{2} \arg (y - y_0) = 0, \pi, 2\pi$$

$$\pi/4 + \tfrac{1}{2} \arg w'(y_0) - \tfrac{1}{2} \arg \nu_1(y_0) + \arg \sigma_0(y_0) + \arg \lambda + \tfrac{3}{2} \arg (y - y_0) = 0, \pi, 2\pi$$

$$\pi/4 + \tfrac{1}{2} \arg w'(y_0) - \tfrac{1}{2} \arg \nu_1(y_0) + \arg (1 \pm \sqrt{\sigma_0(y_0)}) + \arg \lambda + \tfrac{3}{2} \arg (y - y_0) = 0, \pi, 2\pi$$

If $\sigma_0(y)$ does not differ much from a constant and $\operatorname{Im} c$ is small, we find as in the incompressible case [4, p. 131] that the orientation of these curves depends

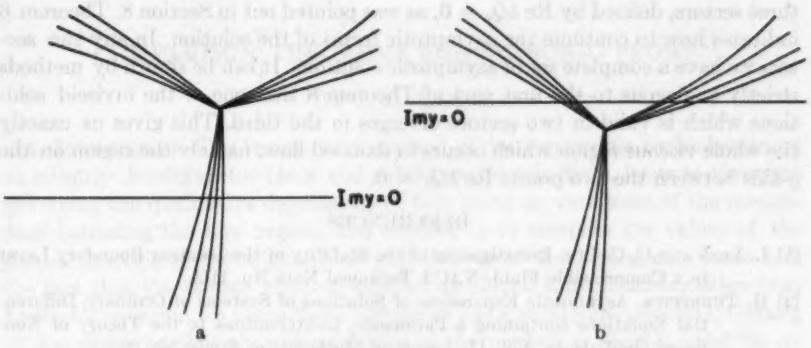



FIG. 5

on the sign of Im $c$. Their position relative to the real $y$-axis is indicated in Figure 5(a) for Im $c > 0$ and in Figure 5(b) for Im $c < 0$, see [1, p. 33]. Now Im $c > 0$ corresponds to unstable solutions while Im $c < 0$ corresponds to damped solutions. The eigen value problem consists in applying boundary conditions at two points on the real axis. In either case, Im $c \gtrless 0$, we may use the fundamental set of solutions of the five lower sectors which contain the end points (Theorem 5). The asymptotic forms are then valid at the end points. If Im $c > 0$ they are valid all along the real axis, i.e., throughout the flow. However, in the case of damped solutions, Figure 5(b), we can use the asymptotic forms only along part of the real axis and their analytic continuation must be determined by a path circling below the origin.

These results proved rigorously here are to be found in [1]. Lin and Lees also shows that if Im $c < 0$ there must exist points on the real axis where the amplitudes of the disturbances depend on the Reynold's number no matter how large the Reynolds number is. This result is corroborated and extended by Theorem 8.

In the hydrodynamic problem, the "power-series" solutions correspond to the inviscid solutions (2.14). In [1, p. 33] it is shown that one of these asymptotic solutions is multivalued near the point $w(y) = c$. By Theorem 5 there exists a corresponding solution for viscous flow which converges in the five lower sectors. This solution can be continued into one more sector up to either Re $\lambda Q_1 = 0$ or Re $\lambda Q_2 = 0$ (Theorem 7). However, it will diverge in at least four of the upper sectors (Theorem 8). In general, then, any stable solution of the flow problem which may be treated as inviscid along most of the real axis will depend very strongly on the Reynold's number either between the two curves Re $\lambda Q_1 = 0$ or the two curves Re $\lambda Q_2 = 0$ which cut the real $y$-axis or possibly between the two outside curves Re $\lambda Q_1 = 0$, Re $\lambda Q_2 = 0$.

Instead of having only four inner viscous layers, [1], we will have a whole section of the flow which is closely dependent on the Reynolds number. The corresponding phenomenon for incompressible flow is contained in [4].

Next we consider the case of constant Prandtl number. Here we have only

three sectors, defined by Re $\lambda Q_1 = 0$, as was pointed out in Section 8. Theorem 6 indicates how to continue the asymptotic forms of the solution. In any two sectors we have a complete set of asymptotic solutions. It can be shown by methods strictly analogous to the first part of Theorem 8 that one of the inviscid solutions which is valid in two sectors diverges in the third. This gives us exactly the whole viscous region which occurs in damped flow, namely the region on the $y$-axis between the two points Re $\lambda Q_1 = 0$.

## REFERENCES

[1] L. Lees and C. C. Lin. Investigation of the Stability of the Laminar Boundary Layer in a Compressible Fluid. NACA Technical Note No. 1115.

[2] H. Turrittin. Asymptotic Expansions of Solutions of Systems of Ordinary Differential Equations containing a Parameter. Contributions to the Theory of Nonlinear Oscillations, Vol. II, Annals of Mathematics Study No. 29.

[3] M. Hukuhara. Sur les propriétés asymptotiques des solutions d'un système d'équations différentielles linéaires contenant un parametre. Memoirs of the Faculty of Engineering, Kyushu Imperial University, Fukuoka, Vol. 8, 1937.

[4] W. Wasow. The Complex Asymptotic Theory of a Fourth Order Differential Equation of Hydrodynamics. Annals of Mathematics, Vol. 49, No. 4, Oct., 1948.

[5] C. C. Lin. On the Stability of Two Dimensional Parallel Flows. Part 1, Quarterly of Applied Mathematics, Vol. 3, 1945.

[6] W. J. Trjitzinsky. Theory of Linear Differential Equations Containing a Parameter, Acta. Mathematica, Vol. 67, 1936.

[7] E. Kamke. Differential Gleichungen Reeler Functionen.

[8] W. Wasow. Asymptotic Solution of the Differential Equation of Hydrodynamic Stability in a Domain containing a Transition Point, Annals of Mathematics, Vol. 58, No. 2, Sept., 1953.

New York University

(Received December 15, 1951. Revised November 19, 1953.)

# COMPARISON THEOREMS FOR SUBSONIC FLOWS*

By J. B. Serrin

**1. Introduction.** The comparison method of incompressible hydrodynamics, as recently developed for plane and axially symmetric flows, leads to theorems governing the qualitative dependence of flow speed on variations of the streamlines bounding the flow region. The method is to compare the values of the stream function of the original flow and the varied flow, using as a basic tool the fact that the stream function satisfies an *elliptic* differential equation (e.g. Laplace's equation for plane flow) and hence has a maximum principle. Gilbarg [1] has observed that the essential idea of the method is equally valid for irrotational subsonic flows of a compressible fluid, since the stream function for such a flow also satisfies an elliptic differential equation. He obtained in this way the following analogue of a result (Lavrentieff's theorem) already known for incompressible flow:

*(A) Let two subsonic flows (plane or axially symmetric) having the same free stream density and velocity be defined in regions $D$ and $\bar{D}$, each bounded by a streamline extending to $\pm\infty$. We suppose that the flow in $D$ or $\bar{D}$ is one of a family of such flows depending continuously on the free stream velocity. Then if $D \subset \bar{D}$ and if the two bounding streamlines have a regular point $P$ in common,[1] the respective flow speeds $q$ and $\bar{q}$ satisfy at $P$ the inequality,*

$$q(P) \leqq \bar{q}(P),$$

*equality holding if and only if the flows are identical.*

Gilbarg pointed out that this theorem may be applied to obtain significant results concerning compressible flows with free boundaries. The following comparison theorem which holds for flows of an incompressible fluid, ([2], [3]), would also have important consequences if extended to subsonic flows of a compressible fluid.

*(B) Under-Over Theorem. Let two flows (plane or axially symmetric) having well-defined free stream velocities be defined in regions $D$ and $\bar{D}$, bounded respectively by streamlines $S$ and $\bar{S}$ extending to $\infty$. Suppose that $S$ and $\bar{S}$ have a common arc $MN$ which forms part of the boundary of a simply connected region $D'$ interior to both $D$ and $\bar{D}$. Let the remaining boundary of $D'$ consist of an arc $NJ$ of $\bar{S}$ and an arc $JM$ of $S$ meeting at a point $J$ common to $S$ and $\bar{S}$; all points of the arc $NJM$ are presumed finite with the possible exception of $J$. Then*

$$\frac{q(M)}{q(N)} \leqq \frac{\bar{q}(M)}{\bar{q}(N)}, \tag{1}$$

* This work supported in part by the Office of Naval Research.

[1] A boundary point of a region is said to be *regular* if it is on the circumference of a circle whose interior falls entirely within the region.

In the original statement of (A) the point $P$ was also not allowed to be on the axis in axially symmetric flow, but this restriction is not necessary (cf. the remark following Lemma 2).

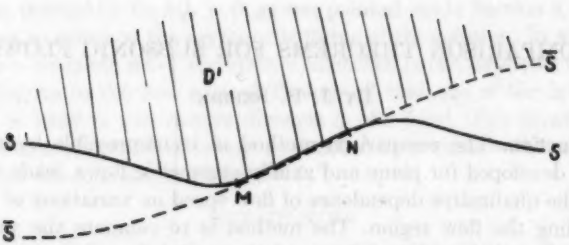



FIG. 1. Configuration illustrating the Under-Over theorem; here $J = \infty$, and both free stream velocities have the same direction.

*equality holding only when the flows are geometrically similar or* $\bar{q}(M) = 0$ *or* $q(N) = 0$.[2]

The purpose of this paper is to make, as far as possible, an extension of this theorem to subsonic flows. Although we do not get a result of equal generality, (even assuming the free stream Mach numbers to be equal), our result does have sufficient validity to be applicable to the compressible free boundary problem, as we shall show in §7, no. 2. Another application concerns the inception of sonic flow on a profile as the free stream Mach number is gradually increased (§7, no. 1).

The basic extension of $(B)$ to subsonic flows is the assertion of its validity *if there exists a point* $P$ *on* $MN$ *where* $q(P) = \bar{q}(P)$. This additional hypothesis could be removed quite simply if, as in the incompressible case, the flows for varying free stream speed (i.e., varying Mach numbers) were dynamically similar, but since this is not true in general, we cannot expect $(B)$ to hold for all subsonic flows. On the other hand, in several special cases the desired similarity does hold: (i) when the flow in $D$ (or $\bar{D}$) is uniform; (ii) when both flows are given by the Prandtl-Glauert approximation

$$(1 - M^2)\psi_{xx} + \psi_{yy} - py^{-1}\psi_y = 0. \qquad M^2 < 1,^3 \tag{2}$$

where (as in the rest of the paper) $p = 0$ or $1$, depending on whether the flow is plane or axially symmetric. In these two cases of subsonic flow, then, $(B)$ holds without restriction. Theorems 1 and 2 of §4 contain the results described in this paragraph; in §5 corresponding theorems are given for flows in a channel.

An important feature of our results is that we permit the points $M$ and $N$ to be on the axis in axially symmetric flow. This is accomplished by generalizing a lemma of E. Hopf [4] concerning the behavior of a solution of an elliptic differential equation at a boundary point. This generalization also allows us to eliminate the condition in $(A)$ that $P$ be excluded from the axis in axially symmetric

[2] $(B)$ was originally stated only for plane flows with a proof depending strongly on a result of function theory [2]. With the additional hypothesis that $M$ and $N$ be regular points not on the axis it was then proved [3] for axially symmetric flows. The additional hypothesis is unnecessary, however, and for this reason it is omitted from the above formulation (cf. the remark concluding §4).

[3] Here the free stream speed is varied without varying $M$, producing a dynamically similar, but artificial, flow to be used for comparison purposes.

flow (see the remark following Lemma 2). Finally in the cases (i) and (ii) above we can dispense entirely with the requirement that $M$ and $N$ be regular boundary points, a condition which seems to be necessary in the more general case.

In [3] I noted some corollaries of $(B)$ arising when the flow in $D$ or $\bar{D}$ is uniform and giving explicit inequalities for the velocity gradient along straight arcs and at inflection points of a streamline. Completely analogous results hold when the flow is subsonic, as we shall show in §6. The paper concludes with a proof of a comparison theorem of a kind different from $(A)$ or $(B)$.

**2. Preliminary results.** We consider an ideal compressible fluid, and flows which are steady, irrotational, and isentropic. Let $\tilde{p}$, $\rho$, and $q$ be the pressure, density, and flow speed; we may assume $q = 1$ for limit speed, $\rho = 0$ for the corresponding density, and a fixed equation of state $\tilde{p} = \tilde{p}(\rho)$. From Bernoulli's law,

$$\int_0^\rho c^2 \, d\rho/\rho = \tfrac{1}{2}(1 - q^2), \qquad c^2 = d\tilde{p}/d\rho > 0, \tag{3}$$

$\rho$ and the sound speed $c$ are obtained as single valued functions of $q$. Introducing the stream function $\psi$ and the mass flow

$$Q = \rho q = y^{-p} \, |\operatorname{grad} \psi| \tag{4}$$

we derive

$$dQ/dq = \rho(1 - m^2), \qquad m^2 = q^2/c^2. \tag{5}$$

Thus as $q$ increases from 0 to sonic speed, $Q$ increases from 0 to some value $b_0$; hence in the totally subsonic range of $Q$, i.e., $0 \leqq Q < b_0$, we have

$$q = q(Q), \qquad \rho = \rho(Q), \qquad c = c(Q). \tag{6}$$

This determined, the stream function of a totally subsonic flow satisfies the quasi-linear elliptic partial differential equation

$$((\rho c)^2 - V^2)\psi_{xx} + 2UV\psi_{xy} + ((\rho c)^2 - U^2)\psi_{yy} - p(\rho c)^2 y^{-1}\psi_y = 0, \tag{7}$$

where $U = y^{-p}\psi_x$, $V = y^{-p}\psi_y$.

If, now, $\psi$ and $\bar{\psi}$ are stream functions for two subsonic flows defined in the same region, say $\Delta$, then in this region $\omega = \bar{\psi} - \psi$ is a solution of a *linear elliptic* partial differential equation *with no term in* $\omega$, namely

$$L(\omega) = A\omega_{xx} + 2B\omega_{xy} + C\omega_{yy} + D\omega_x + (E - pCy^{-1})\omega_y = 0, \tag{8}$$

(see [1], pp. 235–236). Explicit expressions for the coefficients are easily given,

$$A = (\rho c)^2 - V^2 \qquad B = UV, \qquad C = (\rho c)^2 - U^2,$$

$$D = y^{-p}\{(U + \bar{U})(e^2 - \bar{\psi}_{yy}) + (V + \bar{V})\bar{\psi}_{xy}\},$$

$$E = y^{-p}\{(V + \bar{V})(e^2 - \bar{\psi}_{xx}) + (U + \bar{U})\bar{\psi}_{xy}\} - py^{-1}U^2,$$

$$e^2 = \frac{(1 + \bar{\gamma})\bar{\kappa}\bar{m}^2}{2(1 - \bar{m}^2)} \, |\operatorname{grad} \bar{\psi}|, \qquad \gamma = \frac{d \log c^2}{d \log \rho} + 1,$$

where $\kappa$ is the curvature of the streamlines, the bars denote evaluations for the flow given by $\bar{\psi}$, and the tildes denote mean values. The fundamental Lemma 2 below is proved under certain assumptions regarding the boundedness of the coefficients of (8) in the neighborhood of the regular boundary point $P$. In particular, in plane flow, or in axially symmetric flow when $P$ is off the axis, all coefficients must be bounded in the neighborhood of $P$. For this it is sufficient to assume boundedness of the first partial derivatives of the velocity components $\bar{u}$ and $\bar{v}$, or equivalently, boundedness of the acceleration and the curvature (of course, the curvature is bounded *a priori* since $P$ is a regular point). On the other hand, when $P$ is on the axis $A$, $By^{-1}$, $C$, $Dy^{\frac{1}{2}}$, and $E$ must be bounded. This may be insured by requiring boundedness of the first partial derivatives of the velocity components $u$, $v$, $\bar{u}$, and $\bar{v}$, as then the quantities

$$y^{-1}\psi_{xx}\,,\ y^{-1}\psi_{xy}\,,\ \psi_{yy}\,,\ \text{and}\ y^{-1}U$$

will be bounded for both flows. In any event, we assume in all applications of Lemma 2 that the coefficients have the requisite behavior.

**3. Boundary point lemmas.** Since the solutions of an elliptic differential equation of the form (8) satisfy the maximum principle, that is, they cannot have an interior maximum or minimum unless they are constant, we see that the maximum principle holds for the difference $\omega$ of two subsonic stream functions. In addition, under the boundedness conditions just stated, Hopf has shown [4] that if $\omega > 0$ in the neighborhood of a regular point $P$ on the boundary of $\Delta$, while $\omega(P) = 0$, then the inward normal derivative of $\omega$ satisfies the inequality

$$(\partial\omega/\partial n)(P) > 0, \tag{9}$$

provided that $P$ is not on the axis in axially symmetric flow. When $P$ is on the axis we have the corresponding result

$$(\partial\omega/\partial y)(P) = 0, \qquad (\partial^2\omega/\partial y^2)(P) > 0. \tag{9a}$$

This is a consequence of

LEMMA 1. *Let $p$ be a positive constant and let $\omega(x, y)$ be a solution of* (8) *defined in a region $\Delta$ in the upper half plane. Let $P$ be a regular point on the boundary of $\Delta$ with coordinates $(a, 0)$, and suppose that $A$, $By^{-1}$, $C$, $Dy^{\frac{1}{2}}$ and $E$ are bounded and $AC - B^2 \geqq \delta$, $\delta = \text{const.} > 0$, in a neighborhood of $P$. If $\omega > 0$ in this neighborhood and $\omega(P) = 0$ then*

$$\lim_{y\to 0} y^{-p}(\partial\omega/\partial y)(a, y) > 0.$$

*Proof.* We shall assume familiarity with [4]. Let the circle $K$ associated with $P$ have radius $R$, and therefore center at $(a, R)$. We define

$$g(x, y) = 1 - e^{-\alpha(R^2-r^2)},$$

where $\alpha$ is a positive constant to be determined later, and $r^2 = (x - a)^2 + (y - R)^2$. One finds easily that

$$0 < g < 2\alpha R y \qquad \text{in } K.$$

Also put

$$h(x, y) = y^p(e^{-\alpha r^2} - e^{-\alpha R^2});$$

obviously

$$h(a, 0) = 0, \qquad h > 0 \qquad \text{in } K.$$

A simple computation shows that

$$\begin{aligned} e^{\alpha r^2}L(h) &= 2\alpha y^p[2\alpha\{A(x-a)^2 + 2B(x-a)(y-R) + C(y-R)^2\} \\ &\quad - \{A + 2p(x-a)y^{-1}B + (p+1)C + (x-a)D \\ &\quad + E[(pg/2\alpha y) + R - y]\} + \{(pC/y)(R - g/2\alpha y)\}] \\ &> 2\alpha y^p[2\alpha\{A(x-a)^2 + 2B(x-a)(y-R) + C(y-R)^2\} \\ &\quad - \{A + 2p\sqrt{2R}y^{-\frac{1}{2}}|B| + (p+1)C + \sqrt{2R}y^{\frac{1}{2}}|D| \\ &\quad + (p+1)R|E|\}], \end{aligned}$$

the last inequality holding in $K$. In view of the ellipticity condition and the boundedness assumptions we may choose $\alpha$ so large that $L(h) > 0$ in the neighborhood of $P$. It follow in the same way as in [4] that, for suitably small positive numbers $\epsilon$,

$$\omega - \epsilon h > 0$$

in the neighborhood of $P$. Hence

$$\lim_{y\to 0} \frac{1}{y^p}\frac{\partial \omega}{\partial y}(a, y) \geqq \lim_{y\to 0} \frac{\epsilon}{y^p}\frac{\partial h}{\partial y}(a, y) = 2\epsilon\alpha(p+1)\, Re^{-\alpha R^2} > 0. \quad \text{Q.E.D.}$$

We now specialize the preceding remarks to the situations envisaged in theorems $(A)$ and $(B)$. Let $\psi$ and $\bar{\psi}$ denote the respective stream functions, normalized so that

$$\begin{aligned} &\psi = 0 \text{ on } S, \qquad \bar{\psi} = 0 \text{ on } \bar{S}, \\ &\psi > 0 \text{ in } D, \qquad \bar{\psi} > 0 \text{ in } \bar{D}. \end{aligned} \tag{10}$$

The region $\Delta$ becomes the region $D'$ common to $D$ and $\bar{D}$, the boundary of $D'$ then consisting of arcs of $S$ and $\bar{S}$. The function $\omega = \bar{\psi} - \psi$, defined in $D'$, is a solution of (8), and in virtue of the relations (4), (5), (9) and (9a) we get

LEMMA 2. *If $\omega = \bar{\psi} - \psi$ is positive in the neighborhood of a regular boundary point $P$ of $D'$, while $\omega(P) = 0$, then $\bar{q}(P) > q(P)$.*

*Remark.* Since in Lemma 2, $P$ is not excluded from the axis in axially symmetric flow, and since the hypotheses of Lemma 2 are verified in Gilbarg's proof of $(A)$, [1, p. 249], it follows that in the statement of $(A)$ the condition that $P$ be off the axis is unnecessary. For the special cases mentioned in the introduction we have the following improvement of Lemma 2.

LEMMA 3. *If $\omega = \bar{\psi} - \psi$ is positive in the neighborhood of a boundary point $P$ of $D'$, if $\omega(P) = 0$, and if either*

i) *the flow in $\bar{D}$ is one of a family of flows depending continuously on the free stream speed, and $P$ is a regular boundary point of $\bar{D}$, or*

ii) *both flows are given by the Prandtl-Glauert equation* (2), *then* $\bar{q}(P) \geqq q(P)$, *equality holding only in case* ii) *when* $\bar{q}(P) = 0$.

*Proof.* By hypothesis there exists an arc $P_1PP_2$ of the boundary of $D'$ having the property that $\omega > 0$ at all adjacent points in $D'$. Because $\omega \geqq 0$ on $P_1PP_2$ this arc must be on the curve $S$. If $\omega \equiv 0$ on $P_1P$, let $A$ be a regular point of this arc; otherwise let $A$ be a point on $P_1P$ where $\omega > 0$. In the same way choose a point $B$ on $PP_2$, and join $A$ to $B$ by a curve $T$ lying in $D'$, such that $\omega > 0$ on $T$. Let $\bar{\psi}^*$ be the stream function of a flow in $\bar{D}$ with free stream speed $\bar{q}_\infty^* < \bar{q}_\infty$, and set $\chi = \bar{\psi}^* - \bar{\psi}$. Evidently $\chi$ and $\partial\chi/\partial n$ approach zero as $\bar{q}_\infty^* \to \bar{q}_\infty$. Since $\omega > 0$ on $T$, and since at the endpoints of $T$ either $\omega > 0$ or $\omega = 0$ and $\partial\omega/\partial n > 0$,[4] it follows that, for $\bar{q}_\infty^*$ sufficiently near $\bar{q}_\infty$,

$$\bar{\psi}^* - \psi = \omega + \chi > 0 \qquad \text{on } T.$$

Because of (10) and the fact that $P_1PP_2$ is on $S$,

$$\bar{\psi}^* - \psi = \bar{\psi}^* \geqq 0 \qquad \text{on } P_1PP_2,$$

so that, in virtue of the maximum principle, $\bar{\psi}^* - \psi > 0$ in the region bounded by the arcs $T$ and $P_1PP_2$. Hence

$$(\partial\bar{\psi}^*/\partial n)(P) \geqq (\partial\psi/\partial n)(P),$$

or

$$(\partial^2\bar{\psi}^*/\partial y^2)(P) \geqq (\partial^2\psi/\partial y^2)(P)$$

if $P$ is on the axis in axially symmetric flow. From these inequalities and relations (4) and (5) we get

$$\bar{q}^*(P) \geqq q(P).$$

In case i)

$$\bar{q}(P) > \bar{q}^*(P)$$

since $P$ is a regular point of $\bar{D}$;[5] in case ii)

$$\bar{q}(P) = (\bar{q}_\infty/\bar{q}_\infty^*)\bar{q}_\infty^*(P)$$

and the desired conclusion is a consequence of the last three relations.

The following lemma is similar to the preceding two. Although it will not be needed in the present paper it is of interest because it is applicable in some situations where the other two fail, notably when the curvature at $P$ is positively infinite.

LEMMA 4. *If* $\omega = \bar{\psi} - \psi$ *is positive in the neighborhood of a boundary point* $P$ *of* $D'$, *if* $\omega(P) = 0$, *and if* $\bar{\kappa} \geqq 0$ (*or equally* $\kappa \geqq 0$) *in the neighborhood of* $P$, *then* $\bar{q}(P) \geqq q(P)$, *equality holding if and only if* $\bar{q}(P) = 0$.

[4] If the endpoints are on the axis in axial symmetry, this is replaced by $\partial\omega/\partial y = 0$, $\partial^2\omega/\partial y^2 > 0$, but the conclusion remains the same.

[5] See [1], Theorem 2, or note that $\chi < 0$ in $\bar{D}$ and apply Lemma 2, with $\omega = \bar{\psi} - \bar{\psi}^* = -\chi$.

*Proof.* As in Lemma 3 we find that the function

$$\omega^* = (1 - \epsilon)\bar{\psi} - \psi, \qquad \epsilon > 0,$$

is greater than or equal to zero on an arc $P_1PP_2$ of the boundary of $D'$ and greater than zero on a curve $T$ joining $P_1$ and $P_2$, provided that $\epsilon$ is sufficiently small. We shall show that $\omega^*$ cannot possess an interior minimum in the neighborhood of $P$. For it can be shown that $\omega^*$ satisfies the equation

$$L^*(\omega^*) = -(2 - \epsilon)\epsilon\bar{\kappa}\bar{Q}^2 \,|\operatorname{grad}\bar{\psi}|\left\{1 + \frac{(1 + \bar{\gamma})\bar{m}^2}{2(1 - \bar{m}^2)}\right\}, \tag{11}$$

where $L^*$ is a linear elliptic operator of the form (8); in fact the coefficients of $L^*$ are obtained from the coefficients of $L$ by replacing $\bar{\psi}$ with $(1 - \epsilon)\bar{\psi}$. Since the right hand side of (11) is negative by hypothesis $\omega^*$ can have no interior minimum. This being shown, it follows that $\omega^* > 0$ in the region bounded by the arcs $T$ and $P_1PP_2$. Hence

$$(1 - \epsilon)\bar{Q}(P) \geqq Q(P),$$

and from this inequality we derive $\bar{q}(P) > q(P)$, unless $\bar{q}(P) = 0$. Q.E.D.

**4. Under-Over theorem for subsonic flows.** We can now state our main result.

THEOREM 1. *(B) holds for subsonic flows of a compressible fluid, provided that $M$ and $N$ are regular boundary points of $D'$ and*

$$q(P) = \bar{q}(P) \tag{12}$$

*at some regular point $P$ of $MN$. If both flows have the same free stream velocity we suppose additionally that the flow in $D$ or $\bar{D}$ is one of a family of flows depending continuously on the free stream velocity.*[6] *Furthermore under these hypotheses equality can hold in* (1) *only when the flows are identical.*

*Note.* Theorem 1 applies also to *linearized subsonic compressible flow at fixed Mach number $M$* (this of course forces both flows to have the same free stream velocity). Indeed the proof below applies almost word for word since the stream functions $\psi$ and $\bar{\psi}$ of a pair of such flows satisfy the same elliptic differential equation (2). Moreover the final hypothesis of Theorem 1 is unnecessary in this case, since constant multiples of the flow in $D$, say, immediately provide the required family.

*Proof.* When the flows are identical the equality in (1) is trivial, and we have therefore to show that inequality holds in (1) when the flows are not identical. Define $\psi$, $\bar{\psi}$ and $\omega$ as in the remarks preceding Lemma 2. Then in virtue of (10) and the geometrical hypothesis of (B) we have

$$\begin{cases} \omega \geqq 0 & \text{on the arc } JMP \text{ of } S, \\ \omega \leqq 0 & \text{on the arc } PNJ \text{ of } \bar{S}. \end{cases} \tag{13}$$

[6] This hypothesis is verified for plane flows if $D$ or $\bar{D}$ is a region admitting a finite Dirichlet integral. See [1], p. 245, and Shiffman, *On the existence of subsonic flows of a compressible fluid*, J. Rational Mech. Anal., 1, (1952), pp. 605–652; for axially symmetric flows a corresponding result is unknown. It may be remarked that the family in question need only contain flows arbitrarily near the given one in $D$ or $\bar{D}$.

Moreover, a level line $C$, $\omega = 0$, issues from $P$ into $D'$. To see this, note that $P$ must be the limit of a set of points where $\omega = 0$; otherwise $\omega \gtrless 0$ near $P$ which, by Lemma 2, contradicts (12). If these points were isolated then the set of level lines through them would tend to a *limit curve*. On the limit curve, $\omega = \partial\omega/\partial n = 0$, which gives a Cauchy problem for $\omega$ having the unique solution $\omega \equiv 0$. This is impossible, so the points may not be isolated: that is, they form a curve $\omega = 0$ entering $D'$ at $P$.

Now consider the case $J = \infty$. Without loss of generality we may assume the free stream velocities to be $(q_\infty, 0)$ and $(\bar{q}_\infty, \bar{\theta})$, $0 \leqq \bar{\theta} < \pi$, in polar form.

1) $\theta \not\equiv 0$, occurring only for plane flow. Here $D'$ is a *sector like* region with angular opening from $\bar{\theta}$ to $\pi$. We shall construct a curve in $D'$, joining arc $JM$ of $S$ to arc $NJ$ of $\bar{S}$, and containing exactly one point where $\omega = 0$. Define the vectors

$$\mathbf{Q} = (Q_\infty, \pi/2), \qquad \bar{\mathbf{Q}} = (\bar{Q}_\infty, \pi/2 + \bar{\theta}), \qquad \mathbf{W} = \bar{\mathbf{Q}} - \mathbf{Q} = \lim_{x,y\to\infty} \operatorname{grad} \omega.$$

If $\mathbf{W} = (W, \sigma)$ in polar form then obviously $\bar{\theta} + \pi/2 < \sigma < 3\pi/2$. In virtue of this inequality $\mathbf{W}$ has a non-vanishing component in the direction $\bar{\theta}/2$. Now consider the set of straight lines in $D'$ joining $JM$ to $NJ$ and having inclination $\bar{\theta}/2$. If $L$ is any one of these lines sufficiently far from the origin then $\partial\omega/\partial s \neq 0$ on $L$, where $s$ = arc length on $L$. It follows from (13) that $\omega$ must decrease continually on $L$ as we go from $JM$ to $NJ$. Thus $\omega = 0$ at exactly one point $R$ of $L$, as desired. Counting in $C$ any arcs of $JM$ or $NJ$ along its course, it is evident from elementary continuity considerations and the maximum principle that *$C$ joins $P$ to $R$ and has only the point $P$ in common with the arc $MN$*. This defines two (open) subregions of $D'$, say $D_1$ and $D_2$, (see Figure 2). Using (13), we see

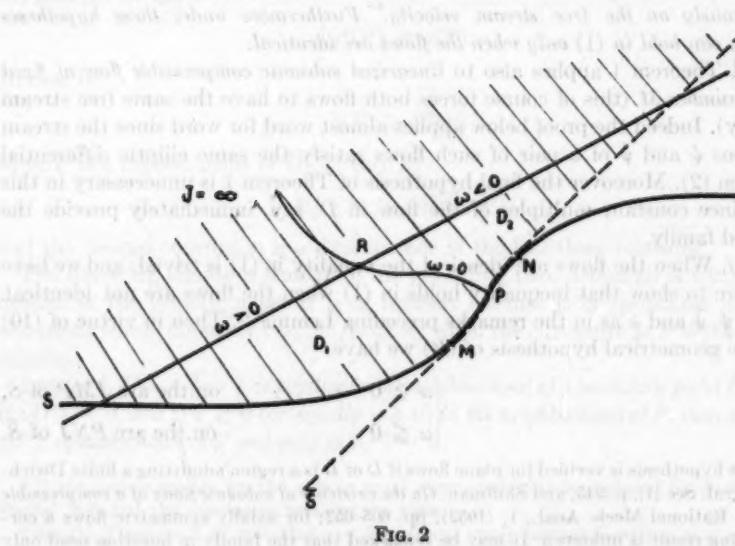



FIG. 2

that $\omega \geqq 0$ on the boundary of $D_1$ and $\omega \leqq 0$ on the boundary of $D_2$, hence according to the maximum principle

$$\omega > 0 \text{ in } D_1, \qquad \omega < 0 \text{ in } D_2$$

Therefore, by Lemma 2,

(14a) $$\bar{q}(M) > q(M),$$

(14b) $$q(N) > \bar{q}(N),$$

from which we derive

$$\frac{q(M)}{\bar{q}(M)} < 1 < \frac{q(N)}{\bar{q}(N)},$$

or

$$\frac{q(M)}{q(N)} < \frac{\bar{q}(M)}{\bar{q}(N)}$$

as desired.

2) $\bar{\theta} = 0$. Here, when $\bar{q}_\infty \neq q_\infty$ we can easily construct a curve $L$ in $D'$, joining $JM$ to $NJ$, and containing at most one change of sign of $\omega$. Following this construction, which we omit, the remainder of the proof is the same as 1). On the other hand, when $\bar{q}_\infty = q_\infty$, corresponding to equal free stream velocities, it is no longer easy to determine a curve $L$ with the required properties. We resort to a more delicate argument.

As before the level line $C$ extends to infinity, and divides $D'$ into two (open) subregions. Let $D_1$ be the subregion bounded by the arc $JMP$ and the level line $C$. We now make use of the assumption that the flow in $D$ ($\bar{D}$ works equally well) is one of a family of flows depending continuously on the free stream speed; let $\psi^*$ denote the stream function for such a flow with free stream speed $q_\infty^* < q_\infty$, and put $\chi = \psi - \psi^*$. Evidently $\chi$ obeys the maximum principle. Therefore, since $\chi = 0$ on $S$ and

$$\lim_{x,y\to\infty} \partial\chi/\partial y > 0$$

we obtain $\chi > 0$ in $D$. Now $\omega \geqq 0$ on the boundary of $D_1$, for this boundary consists of the arc $JMP$ of $S$ and the level line $C$. In view of the two inequalities just written, $\omega + \chi \geqq 0$ on the boundary of $D_1$; moreover because of the equality

$$\omega + \chi = \bar{\psi} - \psi^*$$

we see that $\omega + \chi$ satisfies the maximum principle and

$$\lim_{x,y\to\infty} \frac{\partial(\omega + \chi)}{\partial y} > 0;$$

hence we conclude

$$\omega + \chi > 0 \qquad \text{in } D_1.$$

By making $q_\infty^* \to q_\infty$, so that $\chi \to 0$, we obtain $\omega \geqq 0$ in $D_1$. But $\omega$ is non constant so that application of the maximum principle in its strong form gives $\omega > 0$ in $D_1$. Therefore

$$\bar{q}(M) > q(M) \tag{15a}$$

by Lemma 2. A similar argument leads to

$$q(N) > \bar{q}(N), \tag{15b}$$

and the desired conclusion is a consequence of the last two inequalities. When $J$ is a finite point, a corresponding proof is very easily given since there are no difficulties with growth conditions. We shall therefore go on to the special cases mentioned in the introduction.

THEOREM 2. *(B) holds for subsonic flows of a compressible fluid if the flow in $D$ (or $\bar{D}$) is uniform, or if both flows are defined by the Prandtl-Glauert equation* (2).

*Proof.* In both cases of this theorem the function $\psi^* = a\psi$ satisfies the same equation as $\psi$. By appropriate choice of the constant $a$ we can make

$$q^*(P) = \bar{q}(P)$$

at a regular point $P$ on $MN$ so that we may apply Theorem 1. If $M$ and $N$ are not regular points, Lemma 3 must be used instead of Lemma 2 in obtaining the inequalities (14) and (15) basic to the proof of Theorem 1. We derive in this way

$$\frac{q(M)}{q(N)} = \frac{q^*(M)}{q^*(N)} \leqq \frac{\bar{q}(M)}{\bar{q}(N)} \tag{16}$$

using the fact that the starred and unstarred flows are similar. When the flow in $D$ is uniform, equality can hold in (16) only when the starred and barred flows are identical, or equivalently when the flows in $D$ and $\bar{D}$ are geometrically similar. When both flows are given by the Prandtl-Glauert equation equality can hold only when the flows are identical (geometrical similarity implying identity in this case) or when $\bar{q}(M)$ or $q(N) = 0$.

*Remark.* Since the Prandtl-Glauert theory reduces, when $M = 0$, to incompressible flow theory, Theorem 2 actually contains as a special case the result that $(B)$ holds for flows of an incompressible fluid.

**5. Flows in a channel.** Analogues of Theorems 1 and 2 are available for flows in a channel. We shall be interested specifically in flows bounded by two stream lines having no points in common other than their extremities, which, in particular, may be at infinity; moreover the flow at an extremity shall be uniform or of simple source (or sink) type. If flows of a real gas only are considered it is impossible to define flows of the type described if either extremity is at a finite point. Since it is useful to allow such flow regions for comparison purposes we arrange for them by the following artifice. Let $d = d(Q)$ be defined for all $Q$ so as to agree with $\rho c$ over the range $0 \leqq Q \leqq b_1$, where $b_1$ is any number less than $b_0$, and so that the equation

$$(d^2 - V^2)\psi_{xx} + 2UV\psi_{xy} + (d^2 - U^2)\psi_{yy} - pd^2\frac{\psi_y}{y} = 0 \tag{17}$$

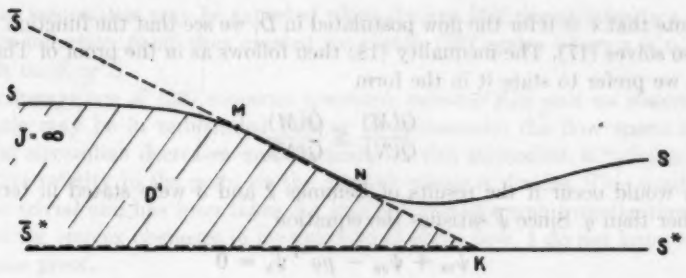



FIG. 3

is elliptic over the whole range $0 \leqq Q < \infty$. Equation (17) can then be used to define a flow in the offending region. Thus, if $\psi$ defines a *bona fide* subsonic flow, if $b_1 = \max Q$, and if $\bar{\psi}$ is the streamfunction of a flow derived from (17), then the developments of §3 hold for the difference $\omega = \bar{\psi} - \psi$ since both $\psi$ and $\bar{\psi}$ satisfy (17). It follows that the methods of §4 may be applied equally well in the envisaged flow situations even though one of the flows may not represent a real gas.

THEOREM 3. *Let two subsonic flows of a compressible fluid be defined in channels $D$ and $\bar{D}$ bounded respectively by streamlines $S$, $S^*$ and $\bar{S}$, $\bar{S}^*$. Let $S$ and $\bar{S}$ have an arc $MN$ in common and suppose the following two conditions are satisfied:*

1. *The curves $S^*$ and $\bar{S}^*$ have a (possibly infinite) arc $JK$ in common, such that $J$ is on $S$ and $K$ is on $\bar{S}$. The arcs $JM$ of $S$ and $NK$ of $\bar{S}$, together with $MN$ and $JK$ bound a simply connected region $JMNK \equiv D'$ interior to $D$ and $\bar{D}$.*[7]

2. *$M$ and $N$ are regular points and there is a regular point $P$ on $MN$ where $q(P) = \bar{q}(P)$. Then*

$$\frac{q(M)}{q(N)} \leqq \frac{\bar{q}(M)}{\bar{q}(N)} \tag{18}$$

*equality holding if and only if the flows are identical.*

THEOREM 4. *If in Theorem 3 the flow in $D$ or $\bar{D}$ is uniform, or if both flows are defined by the Prandtl-Glauert equation, then the restriction 2 need not be made.*

These theorems are the analogues of Theorems 1 and 2 and are proved in the same way (see also [3], p. 568, where some of the details are indicated).

THEOREM 5. *If, in Theorem 3, $\bar{D}$ is a sector and the flow corresponding is of simple source-sink type (Figure 3), then the restriction 2 need not be made, and the conclusion may be written in the form*

$$\frac{\rho q(M)}{\rho q(N)} \leqq \left(\frac{r(N)}{r(M)}\right)^{1+p} \tag{19}$$

*where $r$ is the distance from the vertex $K$ in $\bar{D}$.*

*Proof.* If we write (17) in the form

$$(y^p d)^2(\psi_{xx} + \psi_{yy} - py^{-1}\psi_y) = -\kappa \mid \operatorname{grad} \psi \mid^3$$

[7] It is evident that $J$ is a source (or sink) for the flow in $D$, and $K$ is a source (or sink) for the flow in $\bar{D}$. See Figure 3, where these conditions are fulfilled.

and note that $\kappa \equiv 0$ for the flow postulated in $\bar{D}$, we see that the function $\bar{\psi}^* = a\bar{\psi}$ also solves (17). The inequality (18) then follows as in the proof of Theorem 2 but we prefer to state it in the form

$$\frac{Q(M)}{Q(N)} \leqq \frac{\bar{Q}(M)}{\bar{Q}(N)} \tag{20}$$

which would occur if the results of Lemmas 2 and 3 were stated in terms of $Q$ rather than $q$. Since $\bar{\psi}$ satisfies the equation

$$\psi_{xx} + \psi_{yy} - py^{-1}\psi_y = 0$$

of incompressible fluid theory, we may determine the right hand ratio in (20) from the corresponding incompressible flow. In fact, if $\phi$ is the potential of this flow,

$$\phi = \text{const.} \begin{cases} \log r & \text{plane flow,} \\ 1/r & \text{axially symmetric flow,} \end{cases}$$

then

$$\frac{\bar{Q}(M)}{\bar{Q}(N)} = \frac{|\operatorname{grad} \phi|_M}{|\operatorname{grad} \phi|_N} = \left(\frac{r(N)}{r(M)}\right)^{1+p}. \tag{21}$$

Inequality (19) is now a consequence of (4), (20), and (21).

*Remark.* In this theorem, as well as in any of the preceding ones, the flow in $D$ or $\bar{D}$ need be subsonic only in the subregion $D'$, since the maximum principle has not been applied outside this region. This means that Theorem 5 remains valid in some cases when the flow in the channel $D$ is mixed, as in a transonic nozzle.

**6. Corollaries.** The conclusions of Theorems 2 and 4 may be put in more convenient form. Let $s$ denote arc length along $MN$, increasing from $M$ to $N$. If $P$ is a point on $MN$ with arc length $s$, set $q(P) = q(s)$, $\bar{q}(P) = \bar{q}(s)$. Then for small positive $h$, the conclusions of Theorems 2 or 4 become

$$\frac{q(s)}{q(s+h)} \leqq \frac{\bar{q}(s)}{\bar{q}(s+h)};$$

hence

$$\frac{q(s) - q(s+h)}{hq(s+h)} \leqq \frac{\bar{q}(s) - \bar{q}(s+h)}{h\bar{q}(s+h)},$$

and in the limit $h \to 0$,

$$\frac{d \log q}{ds} \geqq \frac{d \log \bar{q}}{ds}; \tag{22}$$

it follows that

$$\begin{cases} dq/ds \geqq 0 & \text{if the flow in } \bar{D} \text{ is uniform,} \\ d\bar{q}/ds \leqq 0 & \text{if the flow in } D \text{ is uniform.} \end{cases} \tag{23}$$

Similar inequalities may be expected when the arc $MN$ degenerates to a point. At the inflection point thus created we obtain (23) again, where $s$ is now arc length on $S$ or $\bar{S}$.

A consequence of (23) concerns *symmetric subsonic flow past an obstacle* (the obstacle may be in unbounded flow or in a channel): the flow speed on the central streamline decreases monotonically as this streamline is followed from negative infinity to the point on the obstacle where it divides. This result is of course trivial and has been taken for granted in hodograph investigations, but except for convex obstacles in an unbounded plane flow, I do not know of any previous proof.

For Theorem 5 the analogue of (22) is

$$\frac{d \log \rho q}{ds} \geqq \frac{1 + q}{s}$$

where the arc length $s$ is measured from the vertex of $\bar{D}$. If use is made of (3) we get the interesting inequality

$$\frac{d \log q}{ds} \geqq \frac{1 + p}{(1 - m^2)s}.$$

We shall now seek exact solutions of the linearized equation (2) in order to give explicit examples of (22) for this case. Let us consider uniform symmetric flow of Mach number $M < 1$ past a wedge (or cone) of small angle $\alpha$. Taking this for the barred flow will lead to comparison theorems for flows past obstacles like that shaded in Figure 4. Introducing into (2) the new variables

$$\xi = x/\sqrt{1 - M^2}, \qquad \eta = y,$$

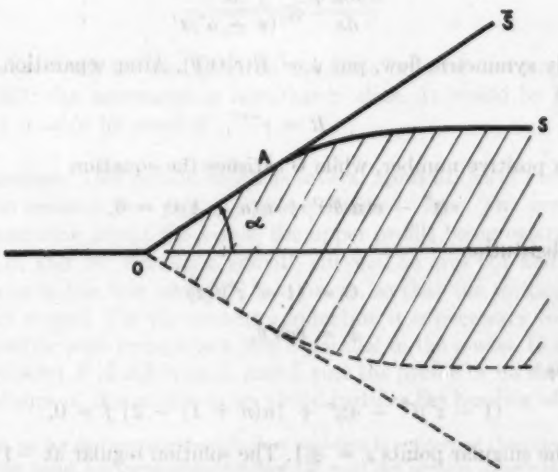



FIG. 4

and polar coordinates $(r, \theta)$ in the $(\xi, \eta)$ plane, (2) becomes

$$\text{(24)} \qquad \begin{cases} \Delta\psi = \psi_{rr} + r^{-2}\psi_{\theta\theta} + r^{-1}\psi = 0 & \text{(plane flow)}, \\ r^2\psi_{rr} + \psi_{\theta\theta} - \operatorname{ctn}\theta\psi_\theta = 0 & \text{(axially symmetric flow)}; \end{cases}$$

appropriate boundary conditions are

$$\text{(24a)} \qquad \begin{cases} \psi(r, \pi) = \psi(r, \alpha') = 0, \\ \tan\alpha' = \sqrt{1 - M^2}\tan\alpha. \end{cases}$$

The solution of (24) for plane flow is obviously

$$\psi = Ar^\beta \sin\beta(\theta - \alpha'),$$

where $\beta = \pi/(\pi - \alpha')$ and $A$ is a constant. For $\theta = \alpha'$,

$$\begin{aligned} \bar{q} &= \sqrt{\psi_x^2 + \psi_y^2} \\ &= [(1 - M^2\cos^2\alpha')/(1 - M^2)]^{\frac{1}{2}}(\psi_\theta/r) \\ &= \text{const. } r^{\beta-1}. \end{aligned}$$

Let $s$ denote arc length on $OA$, measured from $O$. Then since

$$r = \text{const. } s,$$

we have

$$\frac{d\log\bar{q}}{ds} = \frac{\beta - 1}{s}.$$

Thus, by (22), for plane flow past the shaded obstacle

$$\text{(25)} \qquad \frac{d\log q}{ds} \geqq \frac{\alpha'}{(\pi - \alpha')s}, \qquad \text{along } OA.$$

For axially symmetric flow, put $\psi = R(r)\Theta(\theta)$. After separation of variables we find that

$$R = r^{n+1},$$

where $n$ is a positive number, while $\Theta$ satisfies the equation

$$\text{(26)} \qquad \Theta'' - \operatorname{ctn}\theta\Theta' + n(n+1)\Theta = 0.$$

With the substitution

$$\Theta = (1 - z^2)f(z)$$

$$z = \cos\theta,$$

(26) becomes

$$(1 - z^2)f'' - 4zf' + \{n(n+1) - 2\}f = 0,$$

which has the singular points $z = \pm 1$. The solution regular at $-1$ is

$$f(z) = P_{n1}(z) = \frac{dP_n}{dz},$$

TABLE 1

| $\alpha'$ | $n-1$ | $\alpha'(\pi-\alpha')$ |
|---|---|---|
| 0 | 0 | 0 |
| 1.57 | 1 | 1 |
| 2.04 | 2 | 1.85 |
| 2.27 | 3 | 2.61 |
| 2.44 | 4 | 3.48 |
| 2.56 | 5 | 4.41 |
| 2.62 | 6 | 5.10 |
| 3.14 | $\infty$ | $\infty$ |

where $P_n$ is the Legendre function of order $n$; thus

$$\psi = Ar^{n+1} \sin^2 \theta P_{n1}(\cos \theta).$$

To satisfy the boundary conditions we must have

$$P_{n1}(\cos \alpha') = 0$$

$$P_{n1}(\cos \theta) \neq 0, \qquad \alpha' < \theta < \pi.$$

We shall be mainly interested in the value of $n$ corresponding to a given value of $\alpha'$; table 1 presents values of $n-1$ for certain values of $\alpha'$.[5]

Now for $\theta = \alpha'$

$$\bar{q} = y^{-1}\sqrt{\psi_x^2 + \psi_y^2} = \text{const. } r^{n-1},$$

hence inequality (22) becomes, for axially symmetric flow past the shaded obstacle,

$$\frac{d \log q}{ds} \geqq \frac{n-1}{s}. \tag{27}$$

Values of $\alpha'(\pi - \alpha')$ are also given in the table in order to compare formulas (25) and (27); the agreement is remarkably close. It would be interesting to see if $n \approx 1 + \alpha'/\pi$ for small $\alpha'$.

**7. Applications.** This section contains several applications of previous results.

*Symmetric subsonic flow past a profile of "airfoil" type.* The profiles we consider are symmetric about the $x$-axis, the upper profile being composed of three arcs $OA$, $AB$, and $BC$ forming a smooth curve. The arcs $OA$ and $AB$ are presumed convex to the flow while $BC$ is concave, so that the trailing edge or tail $C$ is possibly cusped. For the present application it is necessary to assume that $AB$ is symmetric with respect to a line $L$ parallel to the $y$-axis. It is clear, then, that the midpoint $P$ of $AB$ is on $L$, and $L$ cuts the profile in its thickest section. The resemblance of this profile to an airfoil justifies the heading of this section.

[5] Of course, as far as the compressible flow problem is concerned the interest is in small $\alpha'$. On the other hand, for incompressible flow ($M = 0$) the present theory is exact, and the comparison theorem is useful for large $\alpha = \alpha'$. The table is due to Jahnke-Emde, Funktionentafeln.

Now it is well known that there is a critical free stream Mach number $M_0$ such that the flow is totally subsonic for Mach numbers $M < M_0$, while for $M = M_0$ the sonic speed is reached at one or several points on the (upper) profile. We can prove the following result: *if, in the flow at the critical Mach number, the sonic speed is reached at just one point on the profile,*[9] *say K, then K is forward of P; moreover, if P is on a straight segment $P_1P_2$ of AB, then K is forward of $P_1$.*

For this result the flow may be plane or axially symmetric and may occur in a channel or an unbounded domain. We first prove a lemma: if $M$ is a point on $AP$ and if $N$ is its image in $L$, then in any subsonic flow past the profile the speeds $\bar{q}$ at $M$ and $N$ satisfy the inequality

$$\bar{q}(M) > \bar{q}(N). \tag{28}$$

To prove (28) we apply Theorem 1, taking $\bar{D}$ to be the upper half of the flow past the profile. For $D$ we take the reflection of $\bar{D}$ in $L$, and we let the flow in $D$ be the reflection of the given flow in $\bar{D}$. It is evident that $q(P) = \bar{q}(P)$ so that the hypotheses of Theorem 1 are satisfied;[10] hence

$$\frac{q(M)}{q(N)} < \frac{\bar{q}(M)}{\bar{q}(N)}.$$

But $q(M) = \bar{q}(N)$, $q(N) = \bar{q}(M)$, so that

$$[\bar{q}(M)]^2 > [\bar{q}(N)]^2,$$

and (28) follows.

Now suppose $K$ appeared on $BC$. Consider the flow past the profile at Mach number $M_0$. Since the arc $BC$ is concave the well-known formula

$$\partial q/\partial n = \kappa q$$

shows that the speed increases along the normal to $BC$ at $K$, contradicting the fact that the flow is not supersonic. Suppose next that $K$ appeared on the arc $PB$. Then for $M < M_0$ we have from (28),

$$q(K') > q(K)$$

where $K'$ is the image of $K$ in $L$. In the limit $M \to M_0$ we obtain

$$q(K') = q(K) = c,$$

contradicting the assumption that sonic speed occurs at just one point.

To prove the second part it is sufficient to show that $K$ cannot occur on $P_1P$. Again we shall start with a slightly more general result: in any subsonic flow past the profile the speed $\bar{q}$ is monotonically decreasing along $P_1P$. Consider any two points $M$ and $N$ on $P_1P$ with $N$ nearest $P$. Let $P'$ be the midpoint of $MN$, and let $L'$ be a line parallel to the $y$ axis through $P'$. Finally, let $D$ denote

[9] If this assumption is not made, it can only be asserted that the *set* of sonic points is contained in the arc $OAP$.

[10] In the definition of critical Mach number it is tacitly assumed that the flow in $\bar{D}$ is one of a family of flows depending continuously on the free stream speed.

the reflection of $\bar{D}$ in $L'$. In this case $q(P') = \bar{q}(P')$ so that Theorem 1 may be applied as before, with the result that $\bar{q}(M) > \bar{q}(N)$, i.e. the speed is monotonically decreasing on $P_1P$. That $K$ cannot appear on $P_1P$ now follows in the same way as before.

*Flows with free boundaries.* Comparison theorems $(A)$ and $(B)$ have been used with great success in the study of the classical free boundary problems of hydrodynamics, their application leading specifically to a variety of uniqueness theorems as well as to many qualitative (geometrical) results of importance. The extension of comparison theorems to subsonic flows allows similar conclusions to be drawn for compressible fluids. Indeed, as Gilbarg has observed, the uniqueness of finite cavities and jets carries over to compressible flows with but little alteration in the original proofs, [1]. (The proof of uniqueness for the infinite cavity problem remains incomplete because no formula is available for asymptotic shape of the free boundary, and also for lack of a sufficiently general proof of the comparison theorems.[11])

Likewise Theorems 1 and 2 of reference [5], which are of a geometrical nature, carry over to subsonic flows,[12] although the demonstrations remain incomplete for any axially symmetric problem and for the infinite cavity problem. In the former case this is because the existence of flows in certain auxiliary regions is not known; in the latter, the reason given in footnote 11 stands in the way of a complete proof. Since the required changes in the proofs are not entirely trivial, we shall consider in detail the demonstration of Theorem 1 of [5] for compressible fluids. We shall assume the reader has followed the proof given there through page 565, as no changes are needed up to that point.

The first difficulty encountered is the existence of a flow in the auxiliary region $R_1$. This is assured for plane flow, at least, by the work of Shiffman (see footnote 6); for axially symmetric flow it remains an open question. In any case, for our present purpose we assume that flows in $R_1$ do exist. Let the flow chosen be such that

$$V(P, R_1) = V(P, R) \tag{29}$$

for some point $P$ on $MN$, retaining the notation of [5]. It is convenient to interpret $V(P, R)$ as the value $Q$ at a point $P$ and not as the speed $q$, (thus in the notation of the present paper, (29) would read $Q_1(P) = Q(P)$). By an argument similar to that on page 567, and in virtue of (29), it may be inferred that

$$V(N, R_1) < V(M, R_1). \tag{30}$$

The next step in the proof is to show that $a = (y_m/y_n) < 1$. This result can be obtained as follows. First observe that

$$V(M, R_1) < C/y_m, \tag{31a}$$

[11] That is, the theorems involve embedding a given flow in a family of flows, and it is not known that this is always possible.

[12] Theorem 1 of [5] implies that a starlike obstacle gives rise to a starlike free boundary, a result of importance in the uniqueness problem.

where $C$ is a constant depending only on the strength of the source in $R_1$. This is an immediate consequence of the analogue of $(A)$ for flows in channels, the flow in $R_1$ being compared to a simple source sink flow in a sector. Similarly

$$V(N, R_1) > C/y_N . \tag{31b}$$

Hence

$$V(N, R_1) > a\, V(M, R_1). \tag{32}$$

From (30) and (32) we obtain $a < 1$ as desired. (Note that the original proof does not work since equation (3) is not true for compressible flows.)

The remainder of the proof (p. 568) need not be changed, although it should be pointed out that the existence of a flow in $R_2$ is by no means evident.

*Further applications of comparison methods.* As an application of the methods set out in §4 we consider a type of comparison theorem different from $(A)$ or $(B)$.

*(C) Let a plane subsonic flow having uniform free stream velocity be defined in a region $D$ bounded by a streamline $S$ extending to $\pm\infty$. We suppose that $S$ contains an arc $\sigma$ of a circle whose interior falls entirely outside $D$. Then along the arc $\sigma$ the speed $q$ is first monotonically increasing and then monotonically decreasing.*

*Proof.* To fix the ideas, let the circle have radius $r_0$, let its center be the origin, and take the uniform flow in the positive $x$ direction. Also let the stream function $\psi$ be determined so that $\psi = 0$ on $S$, and $\psi > 0$ in $D$.

Let $r$ designate the radial polar coordinate. For $r \geqq r_0$ define a *circulatory flow* by

$$\bar{q} = kr^{-1}, \qquad \bar{u} = kyr^{-2}, \qquad \bar{v} = -kxr^{-2}, \tag{33}$$

where $k$ is a constant such that $kr_0^{-1} < q^*$, and $q^* = c^* =$ critical speed. The corresponding stream function is given by[13]

$$\bar{\psi} = k \int_{r_0}^{r} (\bar{\rho}/r)\, dr, \tag{34}$$

where $\bar{\rho}$ is given in terms of $r$ by (33) and Bernoulli's law (3). On the boundary $S$ of $D$, $\omega = \bar{\psi} - \psi \geqq 0$, and on $\sigma$, $\omega = 0$. Moreover, $\omega$ has a maximum principle and

$$\lim_{x,y\to\infty} \frac{\partial \omega}{\partial y} < 0. \tag{35}$$

Now let $q_m$ denote the maximum speed on $\sigma$, and let $P$ be any point on $\sigma$ where this speed is attained. If $M$ and $N$ are points of $\sigma$ on one side of $P$, and $N$ is nearest $P$, it is sufficient to show

$$q(M) < q(N) \tag{36}$$

in order to prove $(C)$.

[13] Courant and Friedrichs, *Supersonic Flow and Shock Waves*, N. Y. Interscience (1952), p. 252.

In (33) and (34) put $k = q(N)r_0$; with this choice of $k$ one gets

$$\bar{q} = q(N) = \text{constant} \quad \text{on } \sigma. \tag{37}$$

It follows as in the proof of Theorem 1 that a level line $C$, $\omega = 0$, enters $D$ at $N$. With the help of (35) it is easy to show that $C$ joins $N$ either to $+\infty$ or to $-\infty$. In either case we can apply the usual arguments based on the maximum principle and the boundary point lemmas; depending on which case actually occurs we obtain

$$\bar{q}(P) > q(P) \tag{37}$$

or

$$\bar{q}(M) > q(M). \tag{38}$$

But $\bar{q}(P) = \bar{q}(M) = q(N)$, so that (37) contradicts the fact that the maximum speed is attained at $P$. The remaining inequality (38) then proves (36).

*When the circle falls entirely within $D$*, a slightly refined version of the above proof shows that *along the arc $\sigma$ the speed is first monotonically decreasing and then monotonically increasing*. These theorems remain true in the degenerate case when the circle becomes an entire half plane. In this case the results are also valid for axially symmetric flows.

It is clear that theorems corresponding to $(C)$ exist for incompressible fluids. In fact they take a much more general form due to the *linearity* of the basic equations. The analogue of $(C)$ is: *Let two flows of an incompressible fluid be defined in regions $D$ and $\bar{D}$, bounded respectively by streamlines $S$ and $\bar{S}$. If $\bar{D} \supset D$ and if $S$ and $\bar{S}$ have an arc in common, then along this arc the ratio $q/\bar{q}$ of the speeds is first monotonically increasing and then monotonically decreasing.*

## REFERENCES

[1] D. Gilbarg, *Comparison methods in the theory of subsonic flows*, J. Rational Mech. Anal., 2 (1953), pp. 233–251.

[2] J. Serrin, *Uniqueness theorems for two free boundary problems*, Amer. J. Math., 74 (1952), pp. 492–506.

[3] J. Serrin, *Two hydrodynamical comparison theorems*, J. Rational Mech. Anal., 1 (1952), pp. 563–572.

[4] E. Hopf, *A remark on linear elliptic differential equations of the second order*, Proc. Amer. Math. Soc., 3 (1952), pp. 791–793.

[5] J. Serrin, *On plane and axially symmetric free boundary problems*, J. Rational Mech. Anal., 2 (1953), pp. 563–575.

Massachusetts Institute of Technology

(Received Aug. 4, 1953 (Revised Oct. 9, 1953))

# THE EFFECT OF THICKNESS ON PITCHING AIRFOILS AT SUPERSONIC SPEEDS

By John C. Martin and Nathan Gerber

**Introduction.** The development of the linearized theory of supersonic flow has permitted a first order evaluation of a number of stability derivatives. Second order theories similar to the one introduced by Busemann (reference 1) and extended by Van Dyke (references 2 and 3) offer possibilities of obtaining second order evaluations of certain stability derivatives, such as lift and moment due to steady pitching, and lift and moment due to constant rate of change of angle of attack. The determination of the damping in roll (in reference 4) for certain airfoils is an example of the use of a second order theory to obtain stability derivatives.

In this paper a second order theory is developed for two dimensional pitching airfoils at supersonic speeds. This theory yields an expression for the lifting pressure due to steady pitching which enables the stability derivatives $C_{L_q}$ (lift due to pitching) and $C_{m_q}$ (moment due to pitching) to be calculated. The airfoils considered here have arbitrary symmetrical cross sections; however, the analysis can easily be extended to include airfoils with unsymmetrical cross sections.

The partial differential equation for airfoils with a steady pitching velocity is expressible in a form independent of time. Work by Milton D. Van Dyke (references 2 and 3) indicates that second order solutions of the partial differential equation of steady supersonic flow can be obtained by iterative methods. The equation considered here is quite similar to, though not the same as, the equation of steady supersonic flow. We shall assume that the second order solution can be obtained by iterative methods.

It will also be assumed that the characteristics are the same for the first and the second order solutions (this assumption was made in references 2 and 3). For steady plane flow the second order solution (reference 2) found by using an iterative method based on the preceding assumption yields the correct second order pressure of the Busemann second order theory. Unfortunately, no such justification of this assumption is known to the authors for the flow associated with pitching airfoils.

**The Partial Differential Equation.** The partial differential equation to be used in the following analysis is a special case of the two dimensional time dependent equation for the potential function of a non-viscous compressible fluid. This equation is (see reference 4)

$$a^2(\Phi_{x_0x_0} + \Phi_{z_0z_0}) = \Phi_{tt} + {\Phi_{x_0}}^2\Phi_{x_0x_0} + 2\Phi_{x_0}\Phi_{z_0}\Phi_{x_0z_0} + {\Phi_{z_0}}^2\Phi_{z_0z_0} + 2\Phi_{x_0}\Phi_{x_0t} + 2\Phi_{z_0}\Phi_{z_0t} \tag{1}$$

where $a^2 = C_0^2 - [(\gamma - 1)({\Phi_{x_0}}^2 + {\Phi_{z_0}}^2 + 2\Phi_t)/2]$

This equation is associated with axes fixed in space. For the present problem

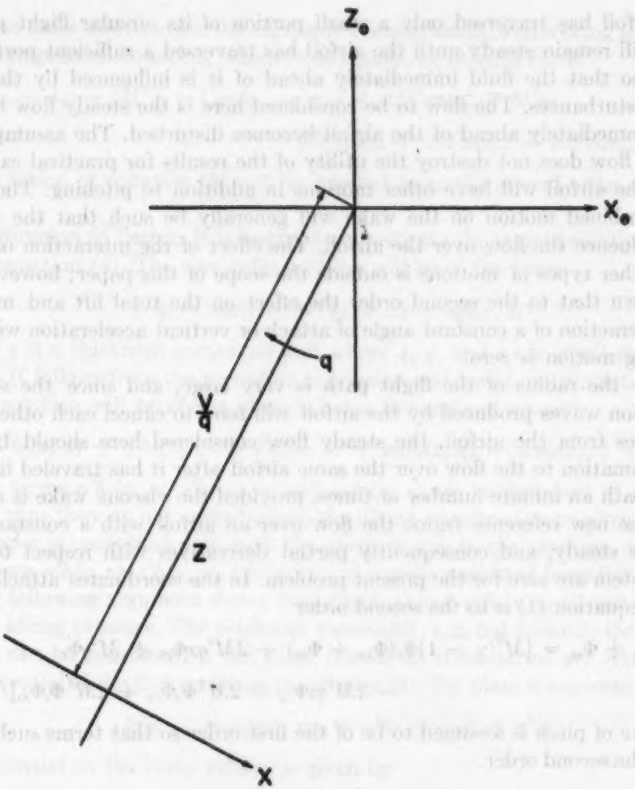



FIG. 1. Stationary and moving axes.

it is convenient to express the differential equation in terms of a set of axes fixed to the airfoil. Since an airfoil having a constant rate of pitch can be considered as flying in a circular flight path with a constant speed, the axes fixed to the airfoil must rotate with respect to the stationary axes with a constant angular velocity $q$. These two sets of axes are illustrated in figure 1. The relations between the two sets of coordinates are

$$x = x_0 \cos qt + z_0 \sin qt \qquad \text{(a)} \qquad (2)$$

$$z = (V/q) - x_0 \sin qt + z_0 \cos qt \qquad \text{(b)}$$

The rates of pitch considered are small so that $qc \ll V$, where $V$ is the velocity and $c$ is the chord of the airfoil. The radius of flight path is $V/q$, and it is seen that this is very large compared to the chord of the airfoil.

We shall assume that the airfoil has traversed only a small arc of its circular flight path so that no effect from its wake is present in the flow over it. Since it is supersonic, the flow over the airfoil will become independent of time after

the airfoil has traversed only a small portion of its circular flight path. The flow will remain steady until the airfoil has traversed a sufficient portion of its circle so that the fluid immediately ahead of it is influenced by the airfoil's own disturbances. The flow to be considered here is the steady flow before the fluid immediately ahead of the airfoil becomes disturbed. The assumption of a steady flow does not destroy the utility of the results for practical calculations since the airfoil will have other motions in addition to pitching. The effect of the combined motion on the wake will generally be such that the wake will not influence the flow over the airfoil. The effect of the interaction of pitching with other types of motions is outside the scope of this paper; however, it can be shown that to the second order the effect on the total lift and moment of the interaction of a constant angle of attack or vertical acceleration with steady pitching motion is zero.

Since the radius of the flight path is very large, and since the shock and expansion waves produced by the airfoil will tend to cancel each other at large distances from the airfoil, the steady flow considered here should be a good approximation to the flow over the same airfoil after it has traveled its circular flight path an infinite number of times, provided the viscous wake is neglected.

In the new reference frame the flow over an airfoil with a constant rate of pitch is steady, and consequently partial derivatives with respect to time in this system are zero for the present problem. In the coordinates attached to the airfoil equation (1) is to the second order

$$-\beta^2\Phi_{xx} + \Phi_{zz} = [M^2(\gamma - 1)\Phi_x(\Phi_{xx} + \Phi_{zz}) + 2M^2qx\Phi_{xz} + M^2q\Phi_z - 2M^2qz\Phi_{xx} + 2M^2\Phi_x\Phi_{xz} + 2M^2\Phi_z\Phi_{zz}]/V \tag{3}$$

The rate of pitch is assumed to be of the first order so that terms such as $qx\Phi_{xz}$ are of the second order.

**The Pressure Relation.** The time dependent pressure relation is (from reference 4)

$$C_p = -\frac{2}{V}(\Phi_{x_0} + \Phi_{t_0}/V) - \frac{1}{V^2}\Phi_{x_0}^2 + \frac{\beta^2}{V^2}\Phi_{z_0}^2 + \frac{2}{V^2}M^2\Phi_{x_0}\Phi_{t_0} + \frac{M^2}{V^4}\Phi_{t_0}^2$$

In the axes attached to the airfoil the preceding equation becomes

$$C_p = -\frac{2}{V}\Phi_x + \frac{\beta^2}{V^2}\Phi_z^2 - \frac{1}{V^2}\Phi_x^2 + \frac{2}{V^2}qz\Phi_x - \frac{2}{V^2}qx\Phi_z \tag{4}$$

**Solution by Iteration.** It will be assumed that equation (3) can be solved by an iteration procedure. The first order solution is taken as the first approximation to the solution. The first order differential equation is obtained by neglecting the second order terms in equation (3), and it is given by

$$-\beta^2\phi_{xx} + \phi_{zz} = 0 \tag{5}$$

It is assumed that the second approximation can be found by substituting

the first order solution into the right side of equation (3) and solving the resulting nonhomogeneous equation, which from equations (3) and (5) is

$$-\beta^2\psi_{xx} + \psi_{zz} = [M^2(\gamma-1)\phi_x(\phi_{xx}+\phi_{zz}) + 2M^2q\phi_{xx} + M^2q\phi_x - 2M^2q\phi_{zz} + 2M^2\phi_x\phi_{xx} + 2M^2\phi_z\phi_{zz}]/V \tag{6}$$

The solution of equation (6) will be referred to as the second order solution.

**Elimination of Terms.** It is helpful to investigate the type of solution obtained from equation (6). The first order solution will be of the form

$$\phi = \alpha g_1(x, z) + qg_2(x, z) + \epsilon g_3(x, z)$$

where $\epsilon$ is a thickness parameter and where $\alpha$, $q$, and $\epsilon$ are small compared to unity. It follows from the preceding expression and equation (6) that the second order solution will be of the form ($\alpha$ being the angle of attack)

$$\psi = \alpha^2h_1(x, z) + q^2h_2(x, z) + \epsilon^2h_3(x, z) + \alpha qh_4(x, z) + \alpha\epsilon h_5(x, z) + q\epsilon h_6(x, z)$$

The terms $\alpha^2h_1(x, z)$, $\epsilon^2h_3(x, z)$, and $\alpha\epsilon h_5(x, z)$ are independent of the rate of pitch; thus they do not contribute to the lifting pressure due to pitching*. Since this paper is concerned only with the pressure due to the pitching, all terms of the forms $\alpha^2h(x, z)$, $\epsilon^2h(x, z)$, and $\alpha\epsilon h(x, z)$ can be neglected hereafter.

The following argument shows that $q^2h_2(x, z)$ and $\alpha qh_4(x, z)$ do not contribute to the lifting pressure. The thickness parameter, $\epsilon$, is not present; therefore, the airfoil can be considered a flat plate insofar as these terms are concerned. If the potential of the flow on the upper surface of a flat plate is expressed as

$$\Phi(\alpha, q) = \alpha g_1 + qg_2 + \alpha^2h_1 + \alpha qh_4 + q^2h_2$$

the potential on the lower surface is given by

$$\Phi(-\alpha, -q) = -\alpha g_1 - qg_2 + \alpha^2h_1 + \alpha qh_4 + q^2h_2$$

The potential difference is

$$\Delta\Phi = 2\alpha g_1 + 2qg_2$$

Since for the flat plate the pressure difference between the upper and lower surfaces can be found directly from the potential difference, the terms $\alpha qh_4$ and $q^2h_2$ do not contribute to the lifting pressure.

The remaining term $qh_6(x, z)$ will be found by the use of equation (6). All expressions multiplied by $\alpha^2$, $\alpha\epsilon$, $q^2$, $\epsilon^2$, and $\alpha q$ will be neglected in the process of evaluating this term.

A further consideration of the form of the solution indicates that the second order lifting pressure is linear in the thickness parameter. This can be established by considering a first order solution of the form

$$\phi = qg_1(x, z) + \epsilon_1g_4(x, z) + \epsilon_2g_5(x, z)$$

* The terms $\alpha^2h_1(x, z)$, and $\alpha\epsilon h_5(x, z)$ are associated with steady two dimensional supersonic flow and can be found by use of equation (46) in reference (2).

The second order lifting pressure will be of the form

$$\Delta C_p = [4qx/(\beta V)] + q\epsilon_1 I_1(x) + q\epsilon_2 I_2(x)$$

This equation is linear in $\epsilon_1$ and $\epsilon_2$ ; thus the lifting pressure for various known distributions can be added to obtain the lifting pressure for new thickness distributions. For the two dimensional airfoil this is of little value for determining analytical solutions, since we shall determine the solution for an arbitrary thickness distribution. This linearity is also true, however, for three dimensional airfoils, and for these it should prove quite useful.

**Boundary Conditions.** Physical considerations require that the flow be tangent to the surface of the airfoil, and that all velocity perturbations vanish upstream of the airfoil.* These boundary conditions may be expressed mathematically as:

$$\phi(x, z) = 0, \qquad \psi(x, z) = 0$$

upstream of the airfoil's leading edge Mach sheet, and

$$\omega \cdot \nabla s = 0$$

where $s(x, z) = 0$ is the equation of the surface of the airfoil.

The equation of the airfoil may also be expressed as

$$z = \epsilon f(x) \tag{7}$$

Thus

$$\nabla s = -i\epsilon \partial f/\partial x + k$$

Since the velocity, $\omega$, can be written as

$$\omega = i(V - zq + \phi_x + \psi_x) + k(qx + \phi_z + \psi_z)$$

it follows that the boundary condition on the body surface is

$$-(V - zq + \phi_x + \psi_x)\epsilon f_x + \phi_z + \psi_z + qx = 0.$$

Thus the boundary conditions for the first order solution are $\phi(x, z) = 0$ upstream of the airfoil's leading edge Mach sheet, and

$$\phi_z|_{z=0} = -qx + V\epsilon f_x \tag{8}$$

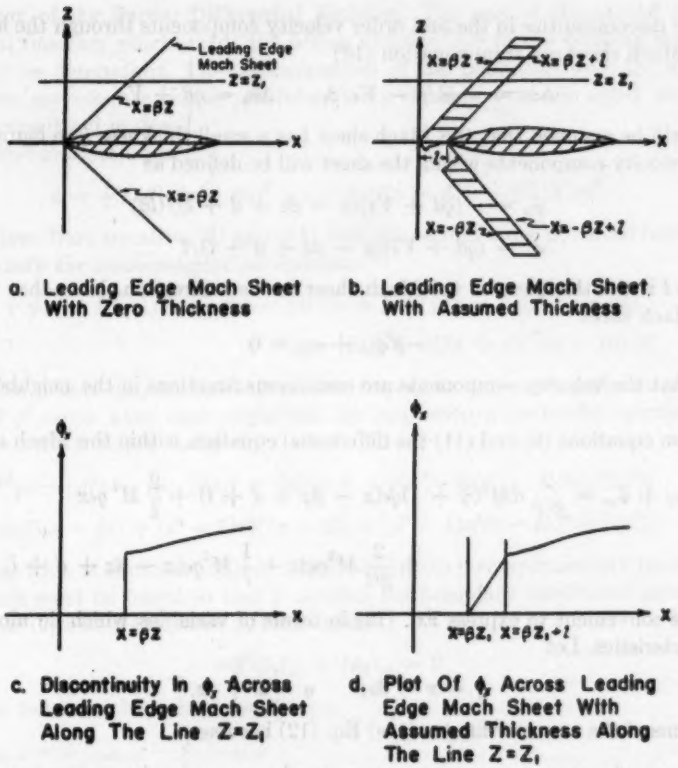
The boundary conditions for the second order solution are $\psi(x, z) = 0$ upstream of the airfoil's leading edge Mach sheet, and

$$\psi_z|_{z=0} = \phi_x|_{z=0}\epsilon f_x - \epsilon f \phi_{zz}|_{z=0} \tag{9}$$

For the airfoils considered here the first order velocity components are discontinuous across the Mach sheet from the leading edge.

To evaluate the effect on the second order solution of the discontinuities in the first order velocity components, we shall assume the Mach sheet from the leading edge to have a small thickness. Within this Mach sheet the first order

* The condition of no disturbances upstream of the airfoil is a result of assuming that the airfoil has traversed only a small arc of the circular flight path.




a. Leading Edge Mach Sheet With Zero Thickness

b. Leading Edge Mach Sheet With Assumed Thickness

c. Discontinuity In $\phi_x$ Across Leading Edge Mach Sheet Along The Line $Z=Z_1$

d. Plot Of $\phi_x$ Across Leading Edge Mach Sheet With Assumed Thickness Along The Line $Z=Z_1$

FIG. 2. An illustration of removing the discontinuity in $\phi_x$ across the leading edge Mach sheet by assuming the Mach sheet to have thickness.

velocity components will be made continuous so that the discontinuities through it are replaced by continuous functions. This process is illustrated in figure 2 for the velocity component in the free stream direction. The effect of the discontinuities will be found by obtaining the second order solution within the Mach sheet and then letting the thickness approach zero.

Only the Mach sheet above the airfoil will be considered since the Mach sheets above and below are of the same form. Since they depend only on the initial slope of the airfoil, the first order velocity discontinuities through the leading edge Mach sheet above the airfoil will be the same as those through the Mach sheet above a flat pitching airfoil with a constant angle of attack of the amount $-\epsilon$.

The first order potential function for the flow over the upper surface of a flat pitching airfoil at an angle of attack $-\epsilon$ is given by

$$\phi = [q(x - \beta z)^2/2 - V\epsilon(x - \beta z) - qd^2/2 - Vcd]/\beta \tag{10}$$

where the airfoil is approximately in the $z = 0$ plane.

The discontinuities in the first order velocity components through the leading edge Mach sheet are from equation (10)

$$\Delta\phi_x = -qd/\beta - V\epsilon/\beta, \qquad \Delta\phi_z = qd + V\epsilon$$

It will be assumed that the Mach sheet has a small thickness (see figure 2-b). The velocity components within the sheet will be defined as

$$\begin{aligned} \phi_x &= -(qd + V\epsilon)(x - \beta z + d + l)/(l\beta) \\ \phi_z &= (qd + V\epsilon)(x - \beta z + d + l)/l \end{aligned} \tag{11}$$

where $l$ is the thickness of the Mach sheet in the $x$ direction. Note that within the Mach sheet

$$-\beta^2\phi_{xx} + \phi_{zz} = 0$$

and that the velocity components are continuous functions in the neighborhood of it.

From equations (6) and (11) the differential equation within the Mach sheet is

$$\begin{aligned} -\beta^2\psi_{xx} + \psi_{zz} = \frac{2}{\beta^2 l^2} dM^4(\gamma + 1)q\epsilon(x - \beta z + d + l) + \frac{2}{l} M^2 q\epsilon x \\ + \frac{2}{\beta^2 l} M^2 q\epsilon\beta z + \frac{1}{l} M^2 q\epsilon(x - \beta z + d + l) \end{aligned} \tag{12}$$

It is convenient to express Eq. (12) in terms of variables which lie along the characteristics. Let

$$\xi = x - \beta z, \qquad \eta = x + \beta z.$$

In terms of the new coordinates $(\xi, \eta)$ Eq. (12) becomes

$$\begin{aligned} \psi_{\xi\eta} = -\frac{d}{2\beta^4 l^2} M^4(\gamma + 1)q\epsilon(\xi + d + l) - \frac{1}{4\beta^2 l} M^2 q\epsilon(\xi + d + l) \\ - \frac{1}{4\beta^2 l} M^2 q\epsilon(\xi + \eta) - \frac{1}{4\beta^4 l} M^2 q\epsilon(\eta - \xi) \end{aligned}$$

This equation can be integrated to yield

$$\begin{aligned} \psi = -\frac{1}{2\beta^4 l^2} dM^4(\gamma + 1)q\epsilon \iint (\xi + d + l)\, d\xi\, d\eta - \frac{1}{4\beta^2 l} M^2 q\epsilon \iint (\xi + \eta)\, d\xi\, d\eta \\ - \frac{1}{4\beta^2 l} M^2 q\epsilon \iint (\xi + d + l)\, d\xi\, d\eta - \frac{1}{4\beta^4 l} M^2 q\epsilon \iint (\eta - \xi)\, d\xi\, d\eta. \end{aligned}$$

The potential on the downstream side of the Mach sheet can be found by evaluating the integrals in the preceding equation and taking the limit as $l \to 0$.

The value of the discontinuity in the second order potential function across the leading edge Mach sheet is (from the above equation)

$$\psi|_{x=\beta z-d} = -\frac{M^2}{2\beta^4} q\epsilon f'(-d)[M^2 x + (2 + \gamma M^2)\, d]\beta z \tag{13}$$

**Solution of the Partial Differential Equation.** The part of the second order potential function which yields the lifting pressure due to the pitching motion will now be determined. The determinations of the potentials over the upper and lower surfaces are similar; therefore, only the flow over the upper surface will be considered in detail.

The first order solution is

$$\phi = \tfrac{1}{2}(1/\beta)q(x - \beta z)^2 - (V/\beta)\epsilon f(x - \beta z) - \tfrac{1}{2}(1/\beta)qd^2 \tag{14}$$

It follows from equation (6) and (14) that the second order potential function must satisfy the nonhomogeneous equation

$$\begin{aligned} -\beta^2\psi_{zz} + \psi_{xx} = M^2q\epsilon\{&f'(x - \beta z) - [M^2(\gamma + 1)/\beta^2][(x - \beta z)f''(x - \beta z) \\ &+ f'(x - \beta z)] + 2(\beta x + z)f''(x - \beta z)/\beta\}, \end{aligned} \tag{15}$$

where the primes denote derivatives with respect to $(x - \beta z)$, and the $\alpha^2$, $\alpha q$, $\alpha\epsilon$, and $q^2$ terms have been neglected. By inspection a particular solution of equation (15) is found to be

$$\begin{aligned} \psi_1 = M^2q\epsilon\{&-[zf(x - \beta z)/(2\beta)] + [M^2(\gamma + 1)z(x - \beta z)f'(x - \beta z)/(2\beta^3)] \\ &+ [-2x\beta zf'(x - \beta z) + (\beta^2 - 1)z^2f'(x - \beta z) + (\beta^2 - 1)zf(x - \beta z)/\beta]/(2\beta^3)\} \end{aligned} \tag{16}$$

The solution of equation (15), $\psi$, consists of $\psi_1$ plus a complementary function, $\psi_2$, which must be found so that $\psi$ satisfies the boundary conditions given by equations (9) and (14). Hence $\psi_2$ must satisfy

$$-\beta^2(\psi_2)_{zz} + (\psi_2)_{xx} = 0$$

and the following boundary conditions:

$$\psi_2 \mid_{x=\beta z-d} = \psi \mid_{x=\beta z-d} - \psi_1 \mid_{x=\beta z-d} \tag{17}$$

$$\begin{aligned} \partial\psi_2/\partial z \mid_{z=0} &= \partial\psi/\partial z \mid_{z=0} - \partial\psi_1/\partial z \mid_{z=0} \\ &= [q\epsilon/(2\beta^3)]\{-M^4\gamma + (\beta^4 + 2\beta^2 - 1)xf'(x) + (1 + \beta^2 - 2\beta^4)f(x)\} \end{aligned} \tag{18}$$

By inspection the complementary function is found to be

$$\begin{aligned} \psi_2 = [q\epsilon/(2\beta^4)]\Big\{&[M^4\gamma - (\beta^4 + 2\beta^2 - 1)](x - \beta z)f(x - \beta z) \\ &+ (-M^4\gamma + 3\beta^4 + \beta^2 - 2)\int_d^{x-\beta z} f(\lambda)\, d\lambda\Big\} \end{aligned} \tag{19}$$

The desired solution is

$$\begin{aligned} \psi = \psi_1 + \psi_2 = [q\epsilon/(2\beta^4)]\Big\{&M^2(M^2\gamma + 1 - \beta^2)\beta xzf'(x - \beta z) \\ &- M^2(M^2\gamma - 2)\beta^2z^2f(x - \beta z) + [-M^2\beta z + M^4\gamma(x - \beta z) \\ &- (\beta^4 + 2\beta^2 - 1)(x - \beta z)]f(x - \beta z) + [-M^4\gamma + 3\beta^4 + \beta^2 - 2]\int_d^{x-\beta z} f(\lambda)\, d\lambda\Big\} \end{aligned} \tag{20}$$

Equation (20) is the part of the second order potential function which yields the entire second order contribution to the lifting pressure due to pitching.

**Lifting Pressure.** The lifting pressure distribution can be expressed as

$$\Delta C_p = -C_p|_{\text{upper surface}} + C_p|_{\text{lower surface}}$$

From equations (4) and (20) it follows that for symmetrical airfoils the lifting pressure distribution is

$$\Delta C_p = -\frac{4q}{\beta V}\left[x + \frac{\epsilon}{2\beta^3}\{M^4\gamma + (\beta^2 - 1)^2\}xf'(x) - \frac{\epsilon}{2\beta^3}M^2 f(x)\right] \quad (21)$$

**Stability Derivatives.** The effect of thickness on the lift and moment, $C_{L_q}$ and

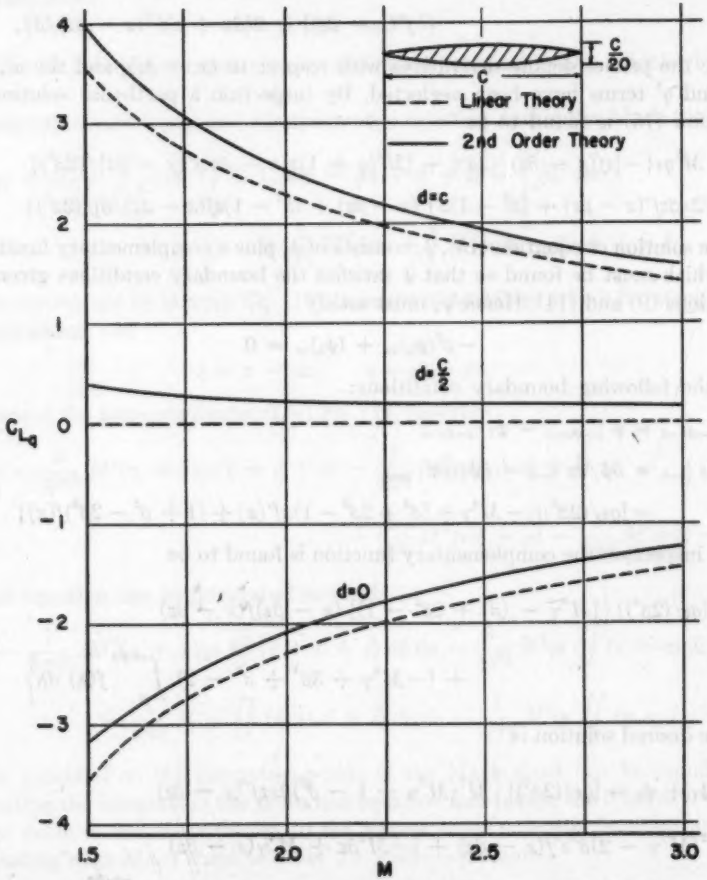



FIG. 3. Variation of the $C_{L_q}$ of a five per cent thick parabolic airfoil with Mach number for various positions of the axis of pitch.

$C_{m_q}$, due to steady pitching can be found by use of equation (21). The stability derivatives $C_{L_q}$ and $C_{m_q}$ can be expressed as

$$C_{L_q} = \left.\frac{\partial C_L}{\partial\left(\frac{qc}{2V}\right)}\right|_{q\to 0} = \frac{1}{\left(\frac{qc^2}{2V}\right)}\int_d^{c-d} \Delta C_p\, dx$$

$$C_{m_q} = \left.\frac{\partial C_m}{\partial\left(\frac{qc}{2V}\right)}\right|_{q\to 0} = \frac{1}{\left(\frac{qc^2}{2V}\right)}\int_d^{c-d} x\Delta C_p\, dx$$

From the preceding relations and equation (21) the $C_{L_q}$ and the $C_{m_q}$ of an airfoil are

$$C_{L_q} = \frac{8d}{\beta c} - \frac{4}{\beta} - \frac{4\epsilon}{\beta^4}\left\{[M^4\gamma + (\beta^2-1)^2]\left[\left(1-\frac{d}{c}\right)\Big/ c\right] f_{TE}\right.$$
$$\left. -[M^4\gamma + (\beta^2-1)^2 + M^2]\int_d^{c-d}\frac{f(\xi)}{c^3}\,d\xi\right\} \qquad (22)$$

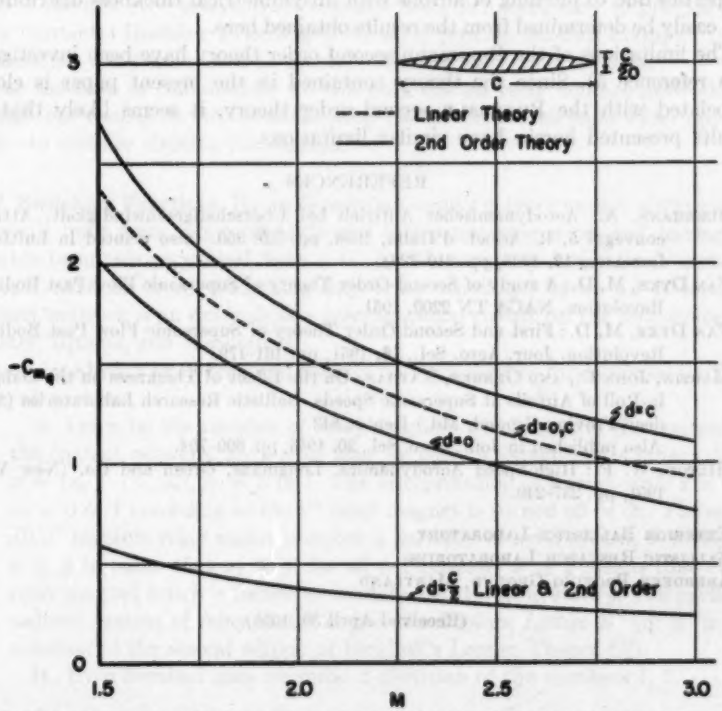



FIG. 4. Variation of the $C_{m_q}$ of a five per cent thick parabolic airfoil with Mach number for various positions of the angle of pitch.

and

$$C_{m_q} = -\frac{8}{3\beta}\left(1 - \frac{3d}{c} + \frac{3d^2}{c^2}\right) - \frac{4\epsilon}{\beta^4}\left\{[M^4\gamma + (\beta^2 - 1)^2]\left[\left(1 - \frac{d}{c}\right)^2 \Big/ c\right] f_{TE} - [2M^4\gamma + 2(\beta^2 - 1)^2 + M^2]\int_d^{c-d} \frac{\xi f(\xi)}{c^3}\, d\xi\right\} \tag{23}$$

Equation (23) indicates that the effect of thickness on the $C_{m_q}$ is zero for an airfoil pitching about the $c/2$ point with a thickness distribution which is chordwise symmetrical with respect to the mid-chord point.

Figure 3 presents the variation of the $C_{L_q}$ for a five per cent parabolic airfoil with Mach number for various positions of the axis of pitch. Figure 4 presents the variation of the $C_{m_q}$ for a five per cent parabolic airfoil with Mach number for various positions of the axis of pitch.

**Concluding Remarks.** The airfoils considered in this paper have symmetrical thickness distributions. But since the flow over the upper and lower surfaces of the airfoils treated here are independent of each other, the aerodynamic properties due to pitching of airfoils with unsymmetrical thickness distributions can easily be determined from the results obtained here.

The limitations of the Busemann second order theory have been investigated (see reference 5). Since the theory contained in the present paper is closely associated with the Busemann second order theory, it seems likely that the results presented herein have similar limitations.

## REFERENCES

1. Busemann, A.: Aerodynamischer Auftrieb bei Überschallgeschwindigkeit, Atti die convegni 5, R. Accad. d'Italia, 1936, pp. 328–360. (also printed in Luftfahrtforschung **12**, 1935, pp. 210–229).
2. Van Dyke, M. D.: A study of Second-Order Theory of Supersonic Flow Past Bodies of Revolution. NACA TN 2200, 1951.
3. Van Dyke, M. D.: First and Second-Order Theory of Supersonic Flow Past Bodies of Revolution. Jour. Aero. Sci., **18**, 1951, pp. 161–179.
4. Martin, John C., and Gerber, Nathan: On the Effect of Thickness on the Damping in Roll of Airfoils at Supersonic Speeds. Ballistic Research Laboratories (Aberdeen Proving Ground, Md.) Report 843.
   Also published in Jour. Aero. Sci., **20**, 1953, pp. 699–704.
5. Hilton, W. F.: High-Speed Aerodynamics. Longmans, Green and Co. (New York) 1951, pp. 237–240.

Exterior Ballistics Laboratory
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

(Received April 30, 1953)

# LATTICE THEORETIC PROPERTIES OF FRONTAL SWITCHING FUNCTIONS

By E. N. Gilbert

**1. Introduction.** It is the aim of this paper to determine the kinds of switching operations which can be performed by switching circuits which use either no negators at all or very few of them. Minimizing the number of appearances of the negation sign (prime) in a Boolean function when the connectives *and* (multiplication), *or* (addition), and *not* (prime) are allowed is a mathematical problem which is fairly closely related to this one.

Practical interest in our problem arises from the fact that in some types of switching circuits, devices which perform negation are more expensive or more unreliable than those which perform other logical operations. For example, in relay contact networks, negation is obtained through the use of back contacts.* The spring adjustments on a relay which has both back and front contacts are more critical than the adjustments on a relay with only front contacts. In the case of the new reed relays, front contacts are much more easy to provide than back contacts. Rectifier switching circuits [1] provide another example; transformers are sometimes used to provide negation in rectifier circuits.

For the sake of concreteness we will only discuss relay contact networks although our results on two terminal relay contact networks have wider application—to rectifier circuits, for example.

**2. Switching Functions.** By an *N-terminal contact network* we will always mean a circuit built from relay contacts and interconnecting wires and having accessible terminals numbered from 1 to $N$. We expressly forbid an $N$-terminal contact network from containing any relay windings; hence an $N$-terminal contact network is an example of a combinational circuit (in the terminology of Keister, Ritchie, and Washburn).

Associated with an $N$-terminal contact network are two fundamental partially ordered systems.

A. Let $n$ be the number of relay magnets which control the contacts in the contact network. By a *relay state* we mean a sequence of $n$ binary digits $x = (x_1, \cdots, x_n)$, $x_i = 0$ or 1. The interpretation of a relay state $x$ is that $x_i = 0$ or 1 according as the $i^{\text{th}}$ relay magnet is turned off or on. The set of all $2^n$ possible relay states becomes a partially ordered system if we define $x \leqq y$ to mean that $x_i \leqq y_i$ for all $i$. Physically $x \leqq y$ means that every relay magnet which is turned on in state $x$ is also on in state $y$. The partially ordered system of relay states is just the *Boolean Lattice* $B^n$ (or $\mathbf{2}^n$ in the notation of the second edition of Birkhoff's Lattice Theory [2]).

B. By a *terminal state* we mean a partition of the numbers $1, 2, \cdots, N$

* Relay contacts are called back (front) contacts if they are closed (open) when the relay magnet is not energized and open (closed) when the magnet is energized.

into any number of indistinguishable groups. Numbers belonging to the same group of a terminal state are meant to represent terminals which are all connected together (electrically) by the contact network. Thus if there are 6 terminals and the network connects terminals 1, 3, and 6 together, 2 and 4 together, and makes no connection to terminal 5 we write the terminal state as (1, 3, 6) (2, 4) (5). The number of terminal states $\Phi(N)$ is not a simple function of $N$ but its generating function [3] is well known:

$$\exp(\exp t - 1) = \sum_{N=0}^{\infty} \frac{\Phi(N)}{N!} t^N.$$

For large $N$ good estimates of $\Phi(N)$ may be obtained from an asymptotic series due to L. F. Epstein [4]. The set of terminal states becomes a familiar partially ordered system, the *partition lattice* $P(N)$, when we define the order relation $\pi_1 \leqq \pi_2$ to mean that state $\pi_1$ is a subpartition of state $\pi_2$. By a subpartition of $\pi_2$ we mean either $\pi_2$ itself or a partition obtainable from $\pi_2$ by breaking down some of the groups appearing in $\pi_2$ into smaller groups; *e.g.*

$$(1)\ (2)\ (3, 6)\ (4)\ (5) \leqq (1, 3, 6)\ (2, 4)\ (5).$$

Physically, $\pi_1 \leqq \pi_2$ here means that every pair of terminals which are connected together in state $\pi_1$ are also connected together in state $\pi_2$.

Given an $N$-terminal contact network, if the relay state is known to be $x$, the terminal state $\pi$ is determined by $x$ and the operation of the network is completely described by giving $\pi$ as a function of $x$. Hence we define an *$N$-terminal switching function* $\pi = f(x)$ to be a mapping of $B^n$ into $P(N)$. Since such a function is specified by associating one of $\Phi(N)$ terminal states with each of the $2^n$ relay states there are

$$(\Phi(N))^{2^n}$$

switching functions. It is an elementary result that every switching function is physically realizable, *i.e.* given any $N$-terminal switching function $f(x)$ one can build an $N$-terminal contact network for which $\pi = f(x)$.

**3. Frontal Switching Functions.** By a *frontal network* we will mean a contact network in which all contacts are front contacts. By a *frontal switching function* we will mean the switching function of a frontal network.

Clearly not all switching functions are frontal switching functions. For example if in relay state $x$ a frontal network connects terminals $a$ and $b$ together, then there is at least one closed path $C_{ab}$ joining $a$ and $b$ in the network. Since $C_{ab}$ contains no back contacts, turning on additional relays (to obtain a relay state $y$ where $x \leqq y$) cannot open $C_{ab}$. Hence every connection made in state $x$ persists in any other state $y$ where $x \leqq y$. In fact, we have

THEOREM 1. *A necessary and sufficient condition that an $N$-terminal switching function $f(x)$ be frontal is that, for all relay states $x$, $y$, if $x \leqq y$ then $f(x) \leqq f(y)$.*

PROOF. We have proved necessity in the preceding paragraph. Sufficiency

will be proved with the aid of some two-terminal frontal networks which we will call $M(\xi)$.

Let $\xi \varepsilon B^n$, say

$$\xi = (\xi_1, \cdots, \xi_n) \qquad \xi_i = 0 \text{ or } 1.$$

The corresponding network $M(\xi)$ is a series connection of front contacts on certain relays. $M(\xi)$ will contain a contact operated by the $i^{th}$ relay if and only if $\xi_i = 1$. The network $M(\xi)$ is a short circuit if $\xi \leqq x$ and an open circuit otherwise.

Let $f(x)$ be a switching function satisfying the condition of Theorem 1. For each pair of terminals $a$, $b$ let $S_{ab}$ be the set of all members $\xi$ of $B^n$ for which $f(\xi)$ requires that $a$ be connected to $b$. Since $S_{ab}$ is itself a partially ordered system it has certain minimal elements $\xi_1, \cdots, \xi_r$. To construct a network for $f(x)$ we need only connect $a$ to $b$ through networks $M(\xi_1), \cdots, M(\xi_r)$ in parallel and repeat this process for every other pair of terminals $a'$, $b'$.

That the network just constructed actually works may be proved as follows:

(i) If terminals $a$ and $b$ are connected in terminal state $f(x)$, then $x \varepsilon S_{ab}$. Hence $\xi_k \leqq x$ for some minimal element $\xi_k$ of $S_{ab}$ and the network $M(\xi_k)$ provides the desired connection.

(ii) If $a$ and $b$ are not connected in state $f(x)$, then $\xi \leqq x$ is false for all elements $\xi$ in $S_{ab}$ (this follows from the condition on $f(x)$ stated in the theorem). Hence none of the direct paths $M(\xi_i)$ joining $a$ and $b$ are closed. The only remaining possible way that $a$ might be joined to $b$ would be by way of a "sneak path" through certain other terminals, say $a$ to $t_1$, $t_1$ to $t_2$, $\cdots$, $t_m$ to $b$. But since terminals $t_i$ and $t_{i+1}$ are connected directly by a network $M(\xi)$ only if they should be connected, this path exists only if $f(x)$ requires $a, t_1, \cdots, t_m, b$ all to be connected together (which is contrary to our assumption).

The above proof of Theorem 1 of course does not completely settle the question of synthesis in practical cases since the network constructed by the process outlined can almost always be much simplified, say by the methods of Boolean algebra [5, 6].

Theorem 1 shows that the notion of frontal switching function is identical with that of isotone (*i.e.* order preserving) mapping from $B^n$ into $P(N)$. We shall need later on in Theorem 5 an order relation in the set of all frontal switching functions. We define $f_1 \leqq f_2$ to hold between two switching functions if $f_1(x) \leqq f_2(x)$ for all $x \varepsilon B^n$. The set of all frontal switching functions $F$ with this order relation is a lattice which, in Birkhoff's notation for cardinal powers, would be written

$$F = (P(N))^{B^n}.$$

**4. Production of Desired Terminal States.** In practice one is often not as much interested in finding a network with a specific switching function $f(x)$ as in finding one which is capable, for some unspecified states of the relay magnets, of interconnecting the terminals in certain desired ways. For example, in designing a telephone switching circuit, one might want a network to connect

the terminals $1, 2, \cdots, N$ together in pairs only; the ways the relays must be set need not be prescribed in advance. The number of relays is not prescribed either. All such problems are solved with $\frac{1}{2}N(N-1)$ relays by joining each pair of terminals through a pair of contacts and operating each pair of contacts by a separate relay. The real problem is to produce the states with a network which minimizes some measure of the network cost, such as the number of relay magnets.

THEOREM 2. *Let $P$ be a partially ordered system composed of certain partitions of $N$ terminals and with the order relation of $P(N)$. In order that a frontal network may exist which produces all the terminal states of $P$ and uses only $n$ relays it is necessary and sufficient that there exist an isomorphism between $P$ and a subset of $B^n$.*

PROOF. If such a network exists, it follows from Theorem 1 that the network's switching function $f(x)$ defines an isomorphism between $P$ and a set of elements of $B^n$ such that $f(x) \varepsilon P$.

Conversely suppose that $P$ consists of terminal states $\pi_1, \cdots, \pi_r$ and let there exist an isomorphism $\pi_i \rightleftarrows X(\pi_i)$ between $P$ and some subset $X(P)$ of $B^n$. By Theorem 1 it suffices to show that there exists an isotone mapping $f(x)$ of $B^n$ into $P(N)$ such that $f(X(\pi_i)) = \pi_i$ for all $\pi_i \varepsilon P$. One such $f(x)$ may be defined as follows:

For each relay state $x \varepsilon B^n$ let $U(x)$ denote the set of all $\pi_i$ of $P$ for which $X(\pi_i) \leqq x$. Since $U(x)$ is a subset of the finite lattice $P(N)$ it follows that l.u.b. $U(x)$ in $P(N)$ exists. Then define

$$f(x) = \text{l.u.b. } U(x).$$

That this definition of $f(x)$ is satisfactory may be proved by noting that

(i) if $x_1 \leqq x_2$, then $U(x_1) \subset U(x_2)$ and $f(x_1) \leqq f(x_2)$,

(ii) $\pi_j \varepsilon U(X(\pi_j))$ and for all $\pi_i$ in $U(X(\pi_j))$, $X(\pi_i) \leqq X(\pi_j)$, or since $X$ is an isomorphism, $\pi_i \leqq \pi_j$; hence $f(X(\pi_j)) = \pi_j$.

In terms of physical networks, the function $f(x)$ constructed by the method of the above proof has the property that, for every terminal pair $a$, $b$, a connection between $a$ and $b$ is made in terminal state $f(x)$ if and only if there is a terminal state $\pi_i \varepsilon P$ in which $a$ and $b$ are connected and such that $X(\pi_i) \leqq x$.

For example, consider the isomorphism between $P \subset P(4)$ and $X(P) \subset B^4$ defined by Table I. The isomorphism as extended to a frontal function $f(x)$ is defined on the rest of $B^4$ by Table II.

In general there will be many other ways of extending a given isomorphism and it may be worth while to search for one which leads to a simple network. For example let $V(x)$ be the set of terminal states $\pi_i$ of $P$ such that $x \leqq X(\pi_i)$; another choice for $f(x)$ could be $f(x) =$ g.l.b. $V(x)$ in $P(N)$.

When $P$ is a complicated system, the problem of finding an isomorphism from $P$ to part of $B^n$ with $n$ as small as possible is difficult enough to make Theorem 2 of limited practical use. It does, however, provide some lower bounds on $n$ even when the isomorphism cannot be found. We obtain this information by considering properties of $P$ which remain invariant under isomorphism such as

(1) the number of elements of $P$

TABLE I

| $\pi_i$ | $X(\pi_i)$ |
|---|---|
| (1) (2) (3) (4) | 0 0 0 0 |
| (12) (3) (4) | 1 1 0 0 |
| (13) (2) (4) | 1 0 1 0 |
| (1) (23) (4) | 0 1 1 0 |
| (14) (2) (3) | 0 1 0 1 |
| (1) (24) (3) | 0 0 1 1 |
| (123) (4) | 1 1 1 0 |
| (1234) | 1 1 1 1 |

TABLE II

| $x$ | $f(x)$ |
|---|---|
| 1 0 0 0 | (1) (2) (3) (4) |
| 0 1 0 0 | (1) (2) (3) (4) |
| 0 0 1 0 | (1) (2) (3) (4) |
| 0 0 0 1 | (1) (2) (3) (4) |
| 1 0 0 1 | (1) (2) (3) (4) |
| 1 1 0 1 | (124) (3) |
| 1 0 1 1 | (13) (24) |
| 0 1 1 1 | (1234) |

(2) the length of the longest chain in $P$ (dimension of $P$)

(3) the largest number of elements of $P$ which can be found which are incomparable with one another.

The invariance of property (1) makes us conclude that $2^n$ must exceed or equal the number of elements of $P$. This is clearly necessary for switching networks of any kind; for networks in which both kinds of contacts are allowed it is sufficient to guarantee that the network sought exists. Some applications of this idea are given by Shannon [7].

Using (2), $n$ must exceed or equal the dimension of $P$.

Using (3), $C_{n,[n/2]}$ must exceed or equal the largest number of incomparable elements which can be found in $P$.*

For example $P(4)$ contains the following set of seven incomparable elements: (123)(4), (124)(3), (134)(2), (1)(234), (12)(34), (13)(24), (14)(23). Hence five relays are required for a frontal network to produce all terminal states of $P(4)$ even though $B^4$ has more elements than $P(4)$.

**5. Two-Terminal Frontal Functions.** $P(2)$ is the lattice of two elements (1)(2), and (12) (*i.e.*, open circuit and closed circuit). To conform to the usual Boolean algebra notation for two-terminal switching functions we agree to let 0 stand for open circuit and 1 for closed circuit; then a two-terminal switching function** becomes a mapping of $B^n$ into the two element lattice $B$ with elements

* That $C_{n,[n/2]}$ is the largest number of incomparable elements of $B^n$ will be shown later.

** In applications to rectifier networks, let 0 stand for ground and 1 for positive; then an

TABLE III

| $n$ | $\psi(n)$ |
|---|---|
| 1 | 3 |
| 2 | 6 |
| 3 | 20 |
| 4 | 168 |
| 5 | 7,581 |
| 6 | 7,828,354 |

0 and 1, $0 \leqq 1$. With this change in notation the notion of two-terminal switching function has its usual significance; *i.e.* as a Boolean function of variables $x_1, \cdots, x_n$ where $x_i \varepsilon B$.

Theorem 1 can be converted into a characterization of frontal two-terminal switching functions in terms of their representations as Boolean functions. In the case of a two-terminal function the synthesis process of Theorem 1 always results in a series-parallel network. Hence we conclude:

REMARK 1. *A Boolean function $f(x_1, \cdots, x_n)$ is a frontal two-terminal switching function if and only if it can be written as an expression containing only the operations addition* (or) *and multiplication* (and). Of course, this observation would have been immediately apparent if we had restricted our attention to series-parallel networks only.

Because of Remark 1, our frontal two-terminal switching functions are precisely the elements of the free distributive lattice $FD(n)$ generated by $n$ symbols $x_1, \cdots, x_n$ (*cf.* Birkhoff, *loc.cit.* pp. 145–146). The problem of counting the number $\psi(n)$ of elements in $FD(n)$ was proposed by Dedekind and has been solved only for $n \leqq 6$. These results are reproduced in Table III. We will use later some crude estimates of $\psi(n)$ which are given by the following theorem.

THEOREM 3. *The number $\psi(n)$ of elements of the free distributive lattice $FD(n)$ (and hence the number of frontal two-terminal switching functions for $n$ relays) satisfies*

$$2^{C_{n,\lfloor n/2 \rfloor}} \leqq \psi(n) \leqq n^{C_{n,\lfloor n/2 \rfloor}} + 2.$$

Proof of lower bound. We may construct $2^{C_{n,\lfloor n/2 \rfloor}}$ different isotone mappings of $B^n$ into $B$ as follows:

(i) Map all sequences with more than $[n/2]$ 0's into 0.

(ii) Map all sequences with more than $n - [n/2]$ 1's into 1.

(iii) The remaining $C_{n,[n/2]}$ elements of $B^n$ (those with exactly $[n/2]$ 0's) can be mapped to 0 or 1 at will and each such assignment completes the definition of an isotone mapping.

This lower bound is about the same as one given by M. Ward [8].

The proof of the upper bound depends on the following lemma.

element $x \varepsilon B^n$ describes the condition of $n$ input leads and $f(x)$ describes the corresponding condition of the output lead of the network.

LEMMA 1. *There exists a set of $C_{n,[n/2]}$ chains in $B^n$, running from $(0, \cdots, 0)$ to $(1, \cdots, 1)$, such that every element of $B^n$ belongs to at least one of the chains.*

The lemma is proved by induction. Let $r \leqq \frac{1}{2}(n-1)$ and suppose $C_{n,r}$ chains have been found running from $(0, \cdots, 0)$ to the $C_{n,r}$ elements of the $r^{\text{th}}$ level (i.e. the elements with $r$ 1's) of $B^n$ and that every element on levels $0, 1, \cdots, r-1$, or $r$ belongs to one of these chains. To each chain we adjoin an element of level $r+1$ in such a way that the same element is never added to more than a single chain.

To show that this is possible we observe that we are concerned with a form of "marriage problem" [9] and hence that a solution exists provided that, for every $\alpha$, every set of $\alpha$ elements on the $r^{\text{th}}$ level are comparable collectively with at least $\alpha$ elements on level $r+1$. Consider the lattice diagram for $B^n$. From a set of $\alpha$ elements on level $r$ there emanate $(n-r)\,\alpha$ lines leading up to some number, say $\beta$, of elements of level $r+1$. These lines are included among the total number $(r+1)\beta$ of lines which lead up to these $\beta$ elements. Thus

$$(n-r)\,\alpha \leqq (r+1)\beta$$

and the marriage problem condition $\alpha \leqq \beta$ is satisfied because $r \leqq \frac{1}{2}(n-1)$.

We now have $C_{n,r}$ chains up to level $r+1$. The remaining $C_{n,r+1} - C_{n,r}$ elements of level $r+1$ can be adjoined in any manner to the original chains up to level $r$ to form chains of level $r+1$. There are then $C_{n,r+1}$ chains to level $r+1$ such that every element of levels $0, 1, \cdots, r$, or $r+1$ belongs to some chain. The induction process can be continued until level $n/2$ (if $n$ is even) or $\frac{1}{2}(n+1)$ (if $n$ is odd) is reached.

Clearly the same process can be used to build chains from $(1, 1, \cdots, 1)$ downward and the chains on the upper part of $B^n$ can be joined to the chains on the lower part of $B^n$ at the level $[n/2]$ to form the $C_{n,[n/2]}$ chains which cover all of $B^n$.

Proof of upper bound. Consider one of the chains of the above lemma. The values of a frontal function $f(x)$ defined on the $n+1$ elements of this chain must form a monotone increasing sequence; thus there are only $n+2$ possible assignments of 0's and 1's to the chain. Of these, two assignments (all 0 or all 1) are exceptional because they are possible only if $f(x)$ is identically 0 or 1. The number of frontal functions, aside from 0 and 1, is certainly less than the number of ways the remaining $n$ assignments can be given to the $C_{n,[n/2]}$ chains (disregarding order relations between elements of different chains) i.e. less than

$$n^{C_{n,[n/2]}}.$$

The bounds given by Theorem 3 differ widely. For example, when $n = 6$ they are about $10^6$ and $3.5 \times 10^9$. Fortunately we are mainly interested in $\log \psi(n)$. For large $n$, we have approximately

$$\sqrt{2/\pi n}\; 2^n \leqq \log \psi(n) \leqq \sqrt{2/\pi n}\, \log_2 n\, 2^n.$$

It is clear that, although there are enormously many frontal functions when $n$ is large, these functions comprise only an infinitesimally small fraction of the total number $2^{2^n}$ of possible switching functions.

Using C. E. Shannon's upper bound $(6K)^K$ on the number of two terminal graphs with $K$ branches, we can readily prove

THEOREM 4. *Let $\epsilon > 0$. The fraction of all frontal two-terminal switching functions which can represent frontal networks with less than*

$$(1 - \epsilon)\frac{\log_2 \psi(n)}{n}$$

*contacts tends to zero as $n \to \infty$.*

Lemma 1 has some other interesting consequences. One of these is the fact, already used in section 4, that no set of incomparable elements of $B^n$ can have more than $C_{n,[n/2]}$ members. Another is an application to punched card sorting which was suggested to the author by E. F. Moore. If we let 0 and 1 stand for hole or no hole in a punched card, an $x$ in $B^n$ represents a way of punching a card with $n$ possible places for holes. If we are to choose a set of ways of punching cards such that when we sort for the cards with punching $x$ we always find $x$ and only $x$, then these ways of punching cards must be represented by a set of incomparable elements of $B^n$. Hence the most ways of punching cards is obtained by choosing the $C_{n,[n/2]}$ ways of punching exactly $[n/2]$ holes.

**6. Non-frontal Two-terminal Switching Functions.** If $f(x)$ is not frontal we would like to represent $f(x)$ in the form

$$f(x) = F(x_1, \cdots, x_n, g_1', \cdots, g_s') \tag{A}$$

where $F$ is a frontal function defined over $B^{n+s}$ and $g_k$ is a frontal function of $(x_1, \cdots, x_n, g_1', \cdots, g_{k-1}')$. It is always possible to do this by taking $s = n$ and $g_i(x) = x_i$. However, we would like to keep $s$ as small as possible.

In the case of rectifier circuits $s$ is the number of negators which would be needed. In the relay case one might construct separate networks for $g_1, \cdots, g_s$ and use these to operate some extra relays containing only back contacts to obtain $g_1', \cdots, g_s'$. Thus $s$ in this case measures the number of extra relays needed. We will show that $s$ need be no greater than about $\log_2 n$. As a first step we convert $f(x)$ into the following normal form.

THEOREM 5. *Given a Boolean function $f(x)$ compute the sequence of functions $\psi_0, \Phi_1, \psi_1, \Phi_2, \cdots$ defined by*

$$\psi_0(x) = f(x)$$

$$\Phi_r(x) = \begin{cases} 1 \text{ if } \psi_{r-1}(y) = 1 \text{ for some } y \leqq x \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

$$\psi_r(x) = \Phi_r(x)\,\psi_{r-1}'(x). \tag{2}$$

*Then the sequence terminates,* i.e. *there is a number $R(f) \leqq n + 1$ such that $\Phi_r(x) = 0$ for all $x$ if $r \geqq R(f)$. Moreover*

$$\Phi_1 > \Phi_2 > \Phi_3 > \cdots \tag{3}$$

*and*

$$f(x) = \Phi_1(x)\{\Phi_2(x)[\Phi_3(x)\ (\cdots)']'\}'. \tag{4}$$

PROOF. The definition (1) of $\Phi_r(x)$ ensures that $\Phi_r(x)$ is a frontal function. The order relation appearing in (3) is the one which was defined for frontal functions in section 3; thus (3) states that $\Phi_1, \Phi_2, \cdots$ form a chain in $FD(n)$. Since no chain in $FD(n)$ is of length greater than $n + 1$, the termination of the process will follow when we prove (3).

Furthermore, by (2),

$$\psi_r + \Phi_r' = \Phi_r\psi_{r-1}' + \Phi_r' = \psi_{r-1}' + \Phi_r'.$$

But by (1) $\psi_{r-1}' + \Phi_r' = \psi_{r-1}'$ so that

$$\psi_{r-1}(x) = \Phi_r(x)\ \psi_r'(x). \tag{5}$$

Using (2) and (5) $r$ times we obtain

$$\begin{aligned} f(x) &= \psi_0(x) \\ &= \Phi_1(x)\ \psi_1'(x) \\ &= \quad \cdot \quad \cdot \\ &\quad \cdot \quad \cdot \quad \cdot \\ &\quad \cdot \quad \cdot \quad \cdot \\ &= \Phi_1(\Phi_2[\Phi_3(\cdots [\Phi_r\ \Psi_r']')']')' \end{aligned}$$

so that (4) will also follow from (3).

To prove (3) let $I_{r-1}$ be the set of elements of $B^n$ for which $\psi_{r-1} = 1$. If $\psi_{r-1}$ is not identically 0 let $x_0$ be a minimal element of $I_{r-1}$ and let $y$ be any element of $B^n$ such that $y < x_0$. By (1), $\Phi_r(y) = 0$ and $\Phi_r(x_0) = 1$. By (2) and (5), $\psi_r(y) = 0$ and $\psi_r(x_0) = 0$. By (1), $\Phi_{r+1}(x_0) = 0$ and $\Phi_{r+1}(y) = 0$. Hence $\Phi_{r+1} < \Phi_r$ if $\psi_{r-1}$ is not identically zero. If $\psi_{r-1}$ is identically zero, $\Phi_r(x) = 0$ by (1) and the sequence has terminated.

Although equation (4) does have the form $(A)$ it is not especially good for our purpose unless the function $\Phi_r = 0$ occurs in the sequence (3) at an early stage. For example, suppose we apply Theorem 5 to the Boolean function

$$f(x) = A^n(x) = \begin{cases} 1 \text{ if an odd number of the variables } x_1, \cdots, x_n \text{ have the value 1,} \\ 0 \text{ otherwise.} \end{cases}$$

We then find that

$$\Phi_r(x) = S_r^n(x) = \begin{cases} 1 \text{ if } r, r+1, \cdots, \text{ or } n \text{ of the variables } x_1, \cdots, x_n \text{ have the} \\ \quad \text{value 1,} \\ 0 \text{ otherwise,} \end{cases}$$

so that $n$ negations are needed in (4).

The synthesis of an arbitrary function $f(x)$ can be reduced to a canonical problem of synthesis of the alternating symmetric function $A^n(x)$.

THEOREM 6. *Let $f(x)$ be a Boolean function for which $\Phi_{m+1} = \Phi_{m+2} = \cdots = \Phi_n = 0$. Then*

$$f(x) = B(\Phi_1, \Phi_2, \cdots, \Phi_m)$$

*where $B(y)$ is the function of $y_1, \cdots, y_n$ such that*

$$A^m = B(S_1^m, S_2^m, \cdots, S_m^m).$$

PROOF. For any given relay state $x \varepsilon B^n$ we have, by (3),

$$\Phi_1 = \Phi_2 = \cdots = \Phi_k = 1$$

$$\Phi_{k+1} = \cdots = \Phi_m = 0$$

for some $k = k(x)$. By (4) we conclude that $f(x) = 1$ if $k$ is odd and $f(x) = 0$ if $k$ is even.

As a simple example of an application of Theorem 6 we observe that

$$A^m = S_1^m(S_2^{m'} + S_3^m)(\cdots)$$

and hence that

$$f(x) = \Phi_1(\Phi_2' + \Phi_3)(\Phi_4' + \Phi_5)\cdots$$

which reduces the number of negators to about $n/2$ at most.

A much more economical synthesis method is obtained through the use of the following representation of $A^{2^s}$ in the form $(A)$ with only $s$ negators. The general case is adequately illustrated by the case of $s = 4$:

$$A^{16} = S_1^{16} g_4' \tag{6}$$

where

$$g_4 = S_2^{16} g_1' g_2' g_3' + S_4^{16} g_1' g_2' + S_6^{16} g_1' g_3' + S_8^{16} g_1' + S_{10}^{16} g_2' g_3' + S_{12}^{16} g_2' + S_{14}^{16} g_3' + S_{16}^{16} \tag{7}$$

where

$$g_3 = S_3^{16} g_1' g_2' + S_7^{16} g_1' + S_{11}^{16} g' + S_{15}^{16} \tag{8}$$

where

$$g_2 = S_5^{16} g_1' + S_{13}^{16} \tag{9}$$

where

$$g_1 = S_9^{16}. \tag{10}$$

The way this representation of $A^{16}$ works may be most simply described by noting that when $k(k > 0)$ of the 16 switching variables have the value 1, the binary representation of $k - 1$ is

$$k - 1 = g_1 2^3 + g_2 2^2 + g_3 2^1 + g_4.$$

Replacing $A^{16}$ by $f$ and $S_r^{16}$ by $\Phi_r$ in (6), $\cdots$, (10) we obtain a representation of type $(A)$ for any Boolean function $f(x)$ for which the chain (3) is of length 16 or less.

Since we can express any Boolean function $f(x)$ with $x \varepsilon B^n$ in the form $(A)$ using no more than about $\log_2 n$ negators it is of interest to see if any further reduction beyond $\log_2 n$ is possible. Actually the best reduction which might be obtained in general can only replace $\log_2 n$ by $K(n) \log_2 n$ where $K(n)$ remains bounded away from 0 as $n \to \infty$. For if every Boolean function of $n$ variables can be expressed in the form $(A)$ with only $s$ negators, $2^{2^n}$ must be less than the number of ways of choosing the $s + 1$ frontal functions of $n + s$ variables $F$, $g_1, \cdots, g_s$:

$$2^{2^n} \leqq [\psi(n + s)]^{s+1}.$$

Using the upper bound on $\psi(n + s)$ derived in Theorem 3 we find that the smallest value $s = s(n)$ which satisfies this inequality is

$$s(n) \approx \tfrac{1}{2} \log_2 n$$

approximately for large $n$.

## REFERENCES

1. W. Keister, A. E. Ritchie and S. H. Washburn, *The Design of Switching Circuits*, Van Nostrand, 1951, p. 218.
2. Garrett Birkhoff, *Lattice Theory*, Am. Math. Soc. Colloquium Publ. Vol. XXV, 1948.
3. W. A. Whitworth, *Choice and Chance*, p. 88, Cambridge, Bell, 1901.
4. L. F. Epstein, J. M. P., **18**, 3, pp. 153–173 (1939).
5. C. E. Shannon, Trans. A. I. E. E., **57**, p. 713, 1938.
6. C. E. Shannon, B. S. T. J., **28**, 1, p. 59.
7. C. E. Shannon, B. S. T. J., **29**, 3, p. 343.
8. M. Ward, Abstract 52-5-135 in Bull Am. Math. Soc. (1946).
9. P. H. Halmos and H. E. Vaughan, Am. J. Math., **72**, 1, p. 214, 1950.

Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey

(Received May 15, 1953)

# A GENERAL SOLUTION OF THE TWO-FREQUENCY MODULATION PRODUCT PROBLEM. II.[1] TABLES OF THE FUNCTIONS $A_{mn}(h, k)$

By Robert L. Sternberg

**1. Introduction.** The purpose of this paper is to provide tables and evaluation methods for the functions

$$(1.1) \qquad A_{mn}(h, k) = (2/\pi^2) \iint_R (\cos u + k \cos v - h) \cos mu du \cos nv dv$$

for $m, n = 0, 1, |h| \leqq 2, 0 < k \leqq 1$ where $R$ is the region defined by

$$(1.2) \qquad R: \cos u + k \cos v \geqq h, \qquad 0 \leqq u, v \leqq \pi,$$

and subsequently also to tabulate briefly the first six higher order functions $A_{mn}(h, k)$. These functions arise in the problem referred to in the title and the theory of their application and a number of their properties have been developed in Part I of this paper by the present writer with H. Kaufman [4] and in the earlier papers of W. R. Bennett [1], [2].

The methods provided for evaluation of the functions (1.1) are indicated in §2 in the description of the construction of the table and consist mainly of special numerical integration schemes based on approximation of the region $R$ defined by (1.2) by suitable polygonal regions. Reference is also made to several known formulas of Bennett [1] for evaluation of the functions (1.1) for $h = 0$ in terms of tabulated complete elliptic integrals and to the double power series expansions given in Part I of the present paper.

The tables provided are six decimal tables of the four basic functions $A_{00}(h, k)$, $A_{10}(h, k)$, $A_{01}(h, k)$ and $A_{11}(h, k)$ for $h = -2(0.2)2$, $k = 0.001, 0.01, 0.1$ and 1.0 and of the first six higher order functions $A_{20}(h, k)$, $A_{02}(h, k)$, $A_{30}(h, k)$, $A_{21}(h, k)$, $A_{12}(h, k)$ and $A_{03}(h, k)$ for $h = -2(0.2)2$ and $k = 0.1$ and 1.0. Each entry in the former set of tables is believed to be accurate to one unit in the last place given while those in the latter set of tables may have errors ranging somewhat larger with, in particular, a maximum error of about three units in the last place being possible in the case of $A_{30}(1.8, 1) = A_{03}(1.8, 1)$.

Viewing the functions $A_{mn}(h, k)$ as new special functions of mathematical physics a name for them seems desirable and, hence, in recognition of the original work of W. R. Bennett [1], [2] in connection with their theory and application, the present writer wishes to take this opportunity to propose that these functions be hereafter known as "Bennett functions."

**2. Construction of the table.** We shall first describe the computation of the four basic functions $A_{00}(h, k)$, $A_{10}(h, k)$, $A_{01}(h, k)$ and $A_{11}(h, k)$.

To begin with for positive values of $h$ in case ($o$) of Part I, i.e., for $h - k \geqq 1$

[1] Presented to the American Mathematical Society at their New York Meeting April 24, 1953.

the tabular entries are trivial and are all zero by formula (4.3) of Part I. For positive values of $h$ not in case ($o$) and $k = 0.001$ and 0.01, and also for $h = 0$, $k = 0.001$, the functions were computed from the approximate formulas

$$
\begin{aligned}
A_{00}(h,k) &\cong (2/\pi^2)\textstyle\sum_{i=0}^{n}[ku_{i+1}\Delta\sin v_i + (\sin u_{i+1} - hu_{i+1})\Delta v_i],\\
A_{10}(h,k) &\cong (2/\pi^2)\textstyle\sum_{i=0}^{n}[k\sin u_{i+1}\Delta\sin v_i\\
&\qquad + (\tfrac{1}{2}u_{i+1} + \tfrac{1}{4}\sin 2u_{i+1} - h\sin u_{i+1})\Delta v_i],\\
A_{01}(h,k) &\cong (2/\pi^2)\textstyle\sum_{i=0}^{n}[ku_{i+1}\Delta(\tfrac{1}{2}v_i + \tfrac{1}{4}\sin 2v_i)\\
&\qquad + (\sin u_{i+1} - hu_{i+1})\Delta\sin v_i],\\
A_{11}(h,k) &\cong (2/\pi^2)\textstyle\sum_{i=0}^{n}[k\sin u_{i+1}\Delta(\tfrac{1}{2}v_i + \tfrac{1}{4}\sin 2v_i)\\
&\qquad + (\tfrac{1}{2}u_{i+1} + \tfrac{1}{4}\sin 2u_{i+1} - h\sin u_{i+1})\Delta\sin v_i],
\end{aligned}
\tag{2.1}
$$

where $(u_0, v_0) = (u_1, 0)$ and the $u_{i+1}, v_{i+1}$, $(i = 0, 1, \cdots, n)$ are the coordinates of suitably chosen points lying near but not necessarily on the curve

$$\cos u + k\cos v = h \tag{2.2}$$

which bounds the region $R$ in (1.1)–(1.2) and where $\Delta v_i \equiv v_{i+1} - v_i$, $\Delta\sin v_i \equiv \sin v_{i+1} - \sin v_i$ and $\Delta(\frac{1}{2}v_i + \frac{1}{4}\sin 2v_i) \equiv \frac{1}{2}v_{i+1} + \frac{1}{4}\sin 2v_{i+1} - \frac{1}{2}v_i - \frac{1}{4}\sin 2v_i$. Formulas (2.1) are readily obtained by approximating the region $R$ in (1.1) – (1.2) by a set of $n + 1$ rectangles having sides parallel to the $u$, $v$-axes with the $(i + 1)$-th rectangle having the vertices

$$(0, v_i),\ (u_{i+1}, v_i),\ (u_{i+1}, v_{i+1}),\ (0, v_{i+1}), \qquad (i = 0, 1, \cdots, n), \tag{2.3}$$

with $(u_{i+1}, v_{i+1})$ located as described. By plotting the curve (2.2) and the above mentioned rectangles accurately together with suitable level curves of the integrand function $\cos u + k\cos v - h$ it was found possible to obtain the desired final accuracy of $1 \times 10^{-6}$ units after round off with the aid of the formulas (2.1) with values of $n$ ranging from one to fourteen for $k = 0.001$ and 0.01.

For positive values of $h$ not in case ($o$) of Part I and $k = 0.1$ the formulas (2.1) are impractical to apply and, hence, in this range the functions were computed from the more suitable approximate formulas

$$
\begin{aligned}
A_{00}(h,k) &\cong \frac{2}{\pi^2}\sum_{i=0}^{n}\left[k\left(\frac{\Delta u_i}{\Delta v_i} + \frac{\Delta v_i}{\Delta u_i}\right)\Delta\cos v_i - \frac{h}{2}(u_i v_{i+1} - u_{i+1}v_i)\right]\\
&\qquad - \frac{h}{\pi}u_{n+1},\\
A_{10}(h,k) &\cong \frac{2}{\pi^2}\sum_{i=0}^{n}\left[-\frac{k}{2}\left(\frac{\Delta v_i}{\Delta u_i + \Delta v_i}\right)\Delta\cos(u_i + v_i)\right.\\
&\qquad - \frac{k}{2}\left(\frac{\Delta v_i}{\Delta u_i - \Delta v_i}\right)\Delta\cos(u_i - v_i) + h\left(\frac{\Delta v_i}{\Delta u_i}\right)\Delta\cos u_i\\
&\qquad \left. - \frac{1}{4}\left(\frac{\Delta v_i}{\Delta u_i}\right)\Delta\cos^2 u_i + \frac{1}{4}(u_i v_{i+1} - u_{i+1}v_i)\right] + \frac{1}{2\pi}u_{n+1},
\end{aligned}
\tag{2.4}
$$

$$A_{01}(h,k) \cong \frac{2}{\pi^2}\sum_{i=0}^{n}\Big[-\frac{1}{2}\Big(\frac{\Delta v_i}{\Delta u_i + \Delta v_i}\Big)\Delta\cos(u_i+v_i) - \frac{1}{2}\Big(\frac{\Delta v_i}{\Delta u_i - \Delta v_i}\Big)\Delta\cos(u_i - v_i) - h\Big(\frac{\Delta u_i}{\Delta v_i}\Big)\Delta\cos v_i + \frac{k}{4}\Big(\frac{\Delta u_i}{\Delta v_i}\Big)\Delta\cos^2 v_i + \frac{k}{4}(u_i v_{i+1} - u_{i+1}v_i)\Big] + \frac{k}{2\pi}u_{n+1},$$

$$A_{11}(h,k) \cong \frac{2}{\pi^2}\sum_{i=0}^{n}\Big[\frac{h}{2}\Big(\frac{\Delta v_i}{\Delta u_i + \Delta v_i}\Big)\Delta\cos(u_i+v_i) + \frac{h}{2}\Big(\frac{\Delta v_i}{\Delta u_i - \Delta v_i}\Big)\Delta\cos(u_i - v_i) - \frac{1}{8}\Big(\frac{\Delta v_i}{2\Delta u_i + \Delta v_i}\Big)\Delta\cos(2u_i+v_i) - \frac{1}{8}\Big(\frac{\Delta v_i}{2\Delta u_i - \Delta v_i}\Big)\Delta\cos(2u_i - v_i) - \frac{k}{4}\Big(\frac{\Delta v_i}{\Delta u_i + 2\Delta v_i}\Big)\Delta\cos(u_i+2v_i) - \frac{k}{4}\Big(\frac{\Delta v_i}{\Delta u_i - 2\Delta v_i}\Big)\Delta\cos(u_i - 2v_i) + \frac{1}{2}\Big(\frac{\Delta u_i}{\Delta v_i} + k^2\frac{\Delta v_i}{\Delta u_i}\Big)\Delta\cos v_i\Big],$$

where $(u_0, v_0) = (\cos^{-1}(h-k), 0)$ and the $u_{i+1}, v_{i+1}$, $(i = 0, 1, \cdots, n)$ now are suitably chosen accurate solutions of the equation (2.2) with, in particular,

$$\begin{aligned}(u_{n+1}, v_{n+1}) &= (0, \cos^{-1}(h' - k')) \quad \text{in case (i), Part I,}\\ (u_{n+1}, v_{n+1}) &= (\cos^{-1}(h+k), \pi) \quad \text{in case (ii), Part I,}\end{aligned} \tag{2.5}$$

where $h' = (h/k), k' = (1/k)$ and as before $\Delta v_i \equiv v_{i+1} - v_i$, $\Delta\cos v_i \equiv \cos v_{i+1} - \cos v_i$ and so forth. Similarly as in the case of (2.1) formulas (2.4) may be derived by approximating the region $R$ in (1.1)–(1.2) by a set of $n+1$ trapezoids having three sides parallel to the $u, v$-axes with the $(i+1)$-th trapezoid having the vertices

$$(0, v_i), (u_i, v_i), (u_{i+1}, v_{i+1}), (0, v_{i+1}), \quad (i = 0, 1, \cdots, n), \tag{2.6}$$

with $(u_i, v_i)$ and $(u_{i+1}, v_{i+1})$ on the curve (2.2). Again by accurately plotting the curve (2.2) and the above mentioned trapezoids together with suitable level curves of the integrand function $\cos u + k\cos v - h$ it was found possible to obtain the desired final accuracy of $1 \times 10^{-6}$ units after round off with the aid of the formulas (2.4) with $n = 7$ for $k = 0.1$.

For positive values of $h$ not in case ($o$) of Part I and $k = 1.0$ the functions $A_{00}(h, 1)$, $A_{10}(h, 1) = A_{01}(h, 1)$ and $A_{11}(h, 1)$ were computed from approximate formulas quite similar to (2.4) but differing in such a manner from those formulas as to take advantage of the symmetry of the region $R$ about the line $u = v$ for $k = 1.0$ to simplify the ultimate calculation.

Finally, for $h = 0$ and $k = 0.01$, 0.1 and 1.0 the functions were computed from the formulas

$$\begin{aligned}&A_{00}(0,1) = 8/\pi^2,\ A_{00}(0,k) = (4/\pi^2)[2E(k) - (1-k^2)K(k)], && k \neq 1,\\ &A_{10}(0,k) = \tfrac{1}{2}, \quad A_{01}(0,k) = \tfrac{1}{2}k, && \text{all } k,\\ &A_{11}(0,1) = 8/3\pi^2,\\ &A_{11}(0,k) = (4/3\pi^2 k)[(1+k^2)E(k) - (1-k^2)K(k)], && k \neq 1,\end{aligned} \tag{2.7}$$

where $K(k)$ and $E(k)$ are respectively the complete elliptic integrals of first and second kind with modulus $k$ as tabulated by Fletcher [3]. These formulas have been given by Bennett [1; pp. 234–235]. For negative values of $h$ the functions $A_{00}(h, k)$, $A_{10}(h, k)$, $A_{01}(h, k)$ and $A_{11}(h, k)$ were computed uniformly by the identities (4.8) of Part I from the already computed values of these functions for positive $h$. This then completed the basic table.

In an effort to ensure accuracy a number of checks were carried out. In particular, for $|h| + k \leqq 0.7$ the tables of the four basic functions were checked to full accuracy by computation with the double power series (5.3), (5.4), (5.5) and (5.6) of Part I using from thirty to forty terms in each series and employing the error estimate (5.7) of Part I or a slightly sharper version of the latter. About half of the remaining entries in the table were checked by recomputation using the formulas (2.1) or (2.4) but employing a different set of points $(u_{i+1}, v_{i+1})$, $(i = 0, 1, \cdots, n)$ from that used in the original computation. The actual arithmetic was performed mainly on a desk calculator and in part on I. B. M. equipment; in either case nine, ten or eleven digits were carried and the final answers were rounded to six decimals.

The computation of the non-trivial values of the higher order functions $A_{20}(h, k)$, $A_{02}(h, k)$, $A_{30}(h, k)$, $A_{21}(h, k)$, $A_{12}(h, k)$ and $A_{03}(h, k)$ was carried out principally with the aid of the recursion identities (4.6) and (4.7) of Part I to evaluate these functions in terms of the four basic functions already tabulated. Thus, in particular, seven place rounded off values of $A_{00}(h, k)$, $A_{10}(h, k)$, $A_{01}(h, k)$ and $A_{11}(h, k)$ were substituted in the formulas (4.7) of Part I and values of the higher order functions were then computed from those formulas to as many significant figures as possible. Although this process resulted in sufficiently accurate values of the higher order functions in most instances, the values of $A_{02}(h, 0.1)$, $A_{03}(h, 0.1)$ and $A_{12}(h, 0.1)$ had to be computed separately due to loss of significant figures in application of the recursion identities. Hence, the non-trivial values of the three functions mentioned with $k = 0.1$ were computed directly from approximate formulas analogous to (2.4) or from the formulas

$$
\begin{aligned}
&A_{20}(0, 1) = A_{02}(0, 1) = 8/9\pi^2, \\
&A_{20}(0, k) = (4/9\pi^2)[2(2 - k^2)E(k) - (1 - k^2)K(k)], && k \neq 1, \\
&A_{02}(0, k) = (4/9\pi^2k^2)[2(2k^2 - 1)E(k) + (2 - 3k^2)(1 - k^2)K(k)], && k \neq 1, \\
&A_{30}(0, k) = A_{21}(0, k) = A_{12}(0, k) = A_{03}(0, k) = 0, && \text{all } k,
\end{aligned} \tag{2.8}
$$

of Bennett [1; p. 234–235]; the latter formulas were also used to partially check the calculations done with the recursion identities. The direct calculation of the functions $A_{02}(h, 0.1)$, $A_{03}(h, 0.1)$ and $A_{12}(h, 0.1)$, like the calculation of the four basic functions, was carried out in such a manner as to yield rounded off six place values accurate to $1 \times 10^{-6}$ units. The somewhat larger possible inaccuracies in the tables of the higher order functions referred to in the introduction occur in the values obtained exclusively by the recursion identities.

**Acknowledgement.** The author wishes to acknowledge his indebtedness to Mrs. S. S. Fahlquist, Mrs. C. F. Ramsey and, particularly, Miss Shirley Rose of

$A_{00}(h, k)$

| $h$ \ $k$ | 0.001 | 0.01 | 0.1 | 1.0 |
|---|---|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .000 000 | .000 000 | .006 476 |
| 1.6 | .000 000 | .000 000 | .000 000 | .026 371 |
| 1.4 | .000 000 | .000 000 | .000 000 | .060 469 |
| 1.2 | .000 000 | .000 000 | .000 000 | .109 678 |
| 1.0 | .000 005 | .000 167 | .005 302 | .175 080 |
| 0.8 | .054 240 | .054 266 | .056 923 | .257 990 |
| 0.6 | .155 095 | .155 115 | .157 090 | .360 067 |
| 0.4 | .288 264 | .288 281 | .290 002 | .483 522 |
| 0.2 | .449 395 | .449 411 | .451 021 | .631 605 |
| 0.0 | .636 620 | .636 636 | .638 212 | .810 569 |
| −0.2 | .849 395 | .849 411 | .851 021 | 1.031 605 |
| −0.4 | 1.088 264 | 1.088 281 | 1.090 002 | 1.283 522 |
| −0.6 | 1.355 095 | 1.355 115 | 1.357 090 | 1.560 067 |
| −0.8 | 1.654 240 | 1.654 266 | 1.656 923 | 1.857 990 |
| −1.0 | 2.000 005 | 2.000 167 | 2.005 302 | 2.175 080 |
| −1.2 | 2.400 000 | 2.400 000 | 2.400 000 | 2.509 678 |
| −1.4 | 2.800 000 | 2.800 000 | 2.800 000 | 2.860 469 |
| −1.6 | 3.200 000 | 3.200 000 | 3.200 000 | 3.226 371 |
| −1.8 | 3.600 000 | 3.600 000 | 3.600 000 | 3.606 476 |
| −2.0 | 4.000 000 | 4.000 000 | 4.000 000 | 4.000 000 |

$A_{10}(h, k)$

| $h$ \ $k$ | 0.001 | 0.01 | 0.1 | 1.0 |
|---|---|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .000 000 | .000 000 | .006 258 |
| 1.6 | .000 000 | .000 000 | .000 000 | .024 582 |
| 1.4 | .000 000 | .000 000 | .000 000 | .054 253 |
| 1.2 | .000 000 | .000 000 | .000 000 | .094 487 |
| 1.0 | .000 005 | .000 167 | .005 215 | .144 416 |
| 0.8 | .052 044 | .052 065 | .054 199 | .203 074 |
| 0.6 | .142 379 | .142 390 | .143 578 | .269 368 |
| 0.4 | .252 316 | .252 323 | .253 012 | .342 032 |
| 0.2 | .373 530 | .373 533 | .373 856 | .419 554 |
| 0.0 | .500 000 | .500 000 | .500 000 | .500 000 |
| −0.2 | .626 470 | .626 467 | .626 144 | .580 446 |
| −0.4 | .747 684 | .747 677 | .746 988 | .657 968 |
| −0.6 | .857 621 | .857 610 | .856 422 | .730 632 |
| −0.8 | .947 956 | .947 935 | .945 801 | .796 926 |
| −1.0 | .999 995 | .999 833 | .994 785 | .855 584 |
| −1.2 | 1.000 000 | 1.000 000 | 1.000 000 | .905 513 |
| −1.4 | 1.000 000 | 1.000 000 | 1.000 000 | .945 747 |
| −1.6 | 1.000 000 | 1.000 000 | 1.000 000 | .975 418 |
| −1.8 | 1.000 000 | 1.000 000 | 1.000 000 | .993 742 |
| −2.0 | 1.000 000 | 1.000 000 | 1.000 000 | 1.000 000 |

$A_{01}(h, k)$

| $h$ \ $k$ | 0.001 | 0.01 | 0.1 | 1.0 |
|---|---|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .000 000 | .000 000 | .006 258 |
| 1.6 | .000 000 | .000 000 | .000 000 | .024 582 |
| 1.4 | .000 000 | .000 000 | .000 000 | .054 253 |
| 1.2 | .000 000 | .000 000 | .000 000 | .094 487 |
| 1.0 | .000 004 | .000 137 | .004 362 | .144 416 |
| 0.8 | .000 205 | .002 048 | .020 329 | .203 074 |
| 0.6 | .000 295 | .002 952 | .029 470 | .269 368 |
| 0.4 | .000 369 | .003 690 | .036 880 | .342 032 |
| 0.2 | .000 436 | .004 359 | .043 582 | .419 554 |
| 0.0 | .000 500 | .005 000 | .050 000 | .500 000 |
| −0.2 | .000 564 | .005 641 | .056 418 | .580 446 |
| −0.4 | .000 631 | .006 310 | .063 120 | .657 968 |
| −0.6 | .000 705 | .007 048 | .070 530 | .730 632 |
| −0.8 | .000 795 | .007 952 | .079 671 | .796 926 |
| −1.0 | .000 996 | .009 863 | .095 638 | .855 584 |
| −1.2 | .001 000 | .010 000 | .100 000 | .905 513 |
| −1.4 | .001 000 | .010 000 | .100 000 | .945 747 |
| −1.6 | .001 000 | .010 000 | .100 000 | .975 418 |
| −1.8 | .001 000 | .010 000 | .100 000 | .993 742 |
| −2.0 | 0.001 000 | 0.010 000 | 0.100 000 | 1.000 000 |

$A_{11}(h, k)$

| $h$ \ $k$ | 0.001 | 0.01 | 0.1 | 1.0 |
|---|---|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .000 000 | .000 000 | .006 046 |
| 1.6 | .000 000 | .000 000 | .000 000 | .022 883 |
| 1.4 | .000 000 | .000 000 | .000 000 | .048 511 |
| 1.2 | .000 000 | .000 000 | .000 000 | .080 846 |
| 1.0 | .000 004 | .000 137 | .004 286 | .117 684 |
| 0.8 | .000 191 | .001 910 | .018 908 | .156 655 |
| 0.6 | .000 255 | .002 546 | .025 387 | .195 121 |
| 0.4 | .000 292 | .002 917 | .029 122 | .229 995 |
| 0.2 | .000 312 | .003 119 | .031 146 | .257 286 |
| 0.0 | .000 318 | .003 183 | .031 791 | .270 190 |
| −0.2 | .000 312 | .003 119 | .031 146 | .257 286 |
| −0.4 | .000 292 | .002 917 | .029 122 | .229 995 |
| −0.6 | .000 255 | .002 546 | .025 387 | .195 121 |
| −0.8 | .000 191 | .001 910 | .018 908 | .156 655 |
| −1.0 | .000 004 | .000 137 | .004 286 | .117 684 |
| −1.2 | .000 000 | .000 000 | .000 000 | .080 846 |
| −1.4 | .000 000 | .000 000 | .000 000 | .048 511 |
| −1.6 | .000 000 | .000 000 | .000 000 | .022 883 |
| −1.8 | .000 000 | .000 000 | .000 000 | .006 046 |
| −2.0 | 0.000 000 | 0.000 000 | 0.000 000 | 0.000 000 |

| $A_{20}(h, k)$ | | | $A_{02}(h, k)$ | | |
|---|---|---|---|---|---|
| $h$ \ $k$ | 0.1 | 1.0 | $h$ \ $k$ | 0.1 | 1.0 |
| 2.0 | 0.000 000 | 0.000 000 | 2.0 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .005 638 | 1.8 | .000 000 | .005 638 |
| 1.6 | .000 000 | .019 756 | 1.6 | .000 000 | .019 756 |
| 1.4 | .000 000 | .038 452 | 1.4 | .000 000 | .038 452 |
| 1.2 | .000 000 | .058 252 | 1.2 | .000 000 | .058 252 |
| 1.0 | .004 958 | .076 181 | 1.0 | .002 275 | .076 181 |
| 0.8 | .046 620 | .089 866 | 0.8 | .001 347 | .089 866 |
| 0.6 | .108 102 | .097 689 | 0.6 | .000 998 | .097 689 |
| 0.4 | .162 196 | .099 053 | 0.4 | .000 870 | .099 053 |
| 0.2 | .198 111 | .094 952 | 0.2 | .000 813 | .094 952 |
| 0.0 | .210 618 | .090 063 | 0.0 | .000 796 | .090 063 |
| −0.2 | .198 111 | .094 952 | −0.2 | .000 813 | .094 952 |
| −0.4 | .162 196 | .099 053 | −0.4 | .000 870 | .099 053 |
| −0.6 | .108 102 | .097 689 | −0.6 | .000 998 | .097 689 |
| −0.8 | .046 620 | .089 866 | −0.8 | .001 347 | .089 866 |
| −1.0 | .004 958 | .076 181 | −1.0 | .002 275 | .076 181 |
| −1.2 | .000 000 | .058 252 | −1.2 | .000 000 | .058 252 |
| −1.4 | .000 000 | .038 452 | −1.4 | .000 000 | .038 452 |
| −1.6 | .000 000 | .019 756 | −1.6 | .000 000 | .019 756 |
| −1.8 | .000 000 | .005 638 | −1.8 | .000 000 | .005 638 |
| −2.0 | 0.000 000 | 0.000 000 | −2.0 | 0.000 000 | 0.000 000 |

$A_{20}(h, k)$

| $h$ \ $k$ | 0.1 | 1.0 |
|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .004 706 |
| 1.6 | .000 000 | .013 303 |
| 1.4 | .000 000 | .019 875 |
| 1.2 | .000 000 | .021 395 |
| 1.0 | .004 552 | .017 339 |
| 0.8 | .035 794 | .009 231 |
| 0.6 | .063 344 | .000 077 |
| 0.4 | .063 717 | −.006 378 |
| 0.2 | .039 001 | −.006 738 |
| 0.0 | .000 000 | .000 000 |
| −0.2 | −.039 001 | .006 738 |
| −0.4 | −.063 717 | .006 378 |
| −0.6 | −.063 344 | −.000 077 |
| −0.8 | −.035 794 | −.009 231 |
| −1.0 | −.004 552 | −.017 339 |
| −1.2 | .000 000 | −.021 395 |
| −1.4 | .000 000 | −.019 875 |
| −1.6 | .000 000 | −.013 303 |
| −1.8 | .000 000 | −.004 706 |
| −2.0 | 0.000 000 | 0.000 000 |

$A_{21}(h, k)$

| $h$ \ $k$ | 0.1 | 1.0 |
|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .005 441 |
| 1.6 | .000 000 | .018 306 |
| 1.4 | .000 000 | .033 958 |
| 1.2 | .000 000 | .048 507 |
| 1.0 | .004 063 | .058 842 |
| 0.8 | .015 018 | .062 662 |
| 0.6 | .015 172 | .058 536 |
| 0.4 | .011 614 | .045 999 |
| 0.2 | .006 213 | .025 729 |
| 0.0 | .000 000 | .000 000 |
| −0.2 | −.006 213 | −.025 729 |
| −0.4 | −.011 614 | −.045 999 |
| −0.6 | −.015 172 | −.058 536 |
| −0.8 | −.015 018 | −.062 662 |
| −1.0 | −.004 063 | −.058 842 |
| −1.2 | .000 000 | −.048 507 |
| −1.4 | .000 000 | −.033 958 |
| −1.6 | .000 000 | −.018 306 |
| −1.8 | .000 000 | −.005 441 |
| −2.0 | 0.000 000 | 0.000 000 |

$A_{12}(h, k)$

| $h$ \ $k$ | 0.1 | 1.0 |
|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .005 441 |
| 1.6 | .000 000 | .018 306 |
| 1.4 | .000 000 | .033 958 |
| 1.2 | .000 000 | .048 507 |
| 1.0 | .002 227 | .058 842 |
| 0.8 | .001 083 | .062 662 |
| 0.6 | .000 601 | .058 536 |
| 0.4 | .000 349 | .045 999 |
| 0.2 | .000 163 | .025 729 |
| 0.0 | .000 000 | .000 000 |
| −0.2 | −.000 163 | −.025 729 |
| −0.4 | −.000 349 | −.045 999 |
| −0.6 | −.000 601 | −.058 536 |
| −0.8 | −.001 083 | −.062 662 |
| −1.0 | −.002 227 | −.058 842 |
| −1.2 | .000 000 | −.048 507 |
| −1.4 | .000 000 | −.033 958 |
| −1.6 | .000 000 | −.018 306 |
| −1.8 | .000 000 | −.005 441 |
| −2.0 | 0.000 000 | 0.000 000 |

$A_{03}(h, k)$

| $h$ \ $k$ | 0.1 | 1.0 |
|---|---|---|
| 2.0 | 0.000 000 | 0.000 000 |
| 1.8 | .000 000 | .004 706 |
| 1.6 | .000 000 | .013 303 |
| 1.4 | .000 000 | .019 875 |
| 1.2 | .000 000 | .021 395 |
| 1.0 | .000 488 | .017 339 |
| 0.8 | −.000 052 | .009 231 |
| 0.6 | −.000 016 | .000 077 |
| 0.4 | −.000 007 | −.006 378 |
| 0.2 | −.000 003 | −.006 738 |
| 0.0 | .000 000 | .000 000 |
| −0.2 | .000 003 | .006 738 |
| −0.4 | .000 007 | .006 378 |
| −0.6 | .000 016 | −.000 077 |
| −0.8 | .000 052 | −.009 231 |
| −1.0 | −.000 488 | −.017 339 |
| −1.2 | .000 000 | −.021 395 |
| −1.4 | .000 000 | −.019 875 |
| −1.6 | .000 000 | −.013 303 |
| −1.8 | .000 000 | −.004 706 |
| −2.0 | 0.000 000 | 0.000 000 |

this laboratory for assistance in performing a portion of the actual calculation and to Mr. R. C. Lesser of the Office of Statistical Services of the Massachusetts Institute of Technology for carrying out an additional part of the calculation on the M. I. T. Model II-I. B. M. Card-Programmed Calculator. In addition the author wishes to acknowledge his indebtedness to Mr. Paul James and Mr. Herbert Johnson for assistance in checking part of the work.

## REFERENCES

1. Bennett, W. R., "New Results in the Calculation of Modulation Products," *Bell System Technical Journal*, vol. 12 (1933), pp. 228–243.
2. Bennett, W. R., "The Biased Ideal Rectifier," *Bell System Technical Journal*, vol. 26 (1947), pp. 139–169.
3. Fletcher, A., "A Table of Complete Elliptic Integrals," *The Philosophical Magazine*, vol. 30 (1940), pp. 516–519.
4. Sternberg, R. L. and Kaufman, H., "A General Solution of the Two-Frequency Modulation Product Problem. I.", *this Journal*, vol. 32 (1953), pp. 233–242.

Laboratory for Electronics, Inc.
Boston, Massachusetts

(Received May 22, 1953)

# ON THE DEGREE OF CONVERGENCE OF SOLUTIONS OF DIFFERENCE EQUATIONS TO THE SOLUTION OF THE DIRICHLET PROBLEM[1]

By J. L. Walsh and David Young

**1. Introduction.** Although finite difference methods are finding frequent use in the solution of boundary value problems associated with partial differential equations of elliptic type, little appears to be known about the accuracy of these methods. The object of the present paper is to study degree of convergence of the difference equation solution of Laplace's equation in the square, as the mesh size approaches zero, under various continuity hypotheses on the boundary values. Gerschgorin [2][2] has given some error bounds but these involve bounds on certain partial derivatives of the solution in the closed region. Finite bounds for these derivatives may not exist and even when they do exist are very hard to find except under very stringent assumptions on the boundary values. The only available error bounds not involving higher derivatives of the solution are the non-explicit bound given by Rosenbloom [10] and the explicit error bounds given by Wasow [13] and by Walsh and Young [12] for the case of the rectangle. In [13] very severe differentiability conditions were imposed. In [10] and [12] uniform error bounds were obtained which are, in general, much too large for a particular point. For instance in [12], the uniform error was shown to vanish of the order $h^{2/7}$, where $h$ is the mesh size, provided the boundary values satisfy a Lipschitz condition of order one. For this case we now show that the error tends to zero as the first power of $h$ uniformly on any closed interior region.

We also obtain other results on the degree of convergence uniformly in any closed interior region of the square expressed in terms of assumed modulus of continuity of the given function, or modulus of continuity except for finite jumps, or modulus of continuity of the derivative, or total variation.

Our results can be easily extended to rectangles, and it is hoped that they will give some indication as to the situation to be expected for more general regions.

**2. Explicit solutions.** As general background for our treatment we state

Theorem 2.1. *Let $R$ denote a bounded plane region whose boundary $S$ consists of a finite number of mutually disjoint Jordan curves, and let the function $F(x, y)$ be bounded on $S$ and continuous on $S$ except perhaps for a finite number of points $Q$. Then there exists a unique function $u(x, y)$ harmonic and bounded in $R$, and continuous in $\Omega = R + S$ except at the points $Q$, and equal to $F(x, y)$ on $S$.*

*If the functions $u_k(x, y)$ harmonic in $R$ are bounded in their totality, if each function is continuous in $\Omega$ except perhaps in the points $Q$, and if $u_k(x, y)$ approaches*

[1] This research was supported, in part, by the Office of Ordnance Research, U. S. Army, under Contract DA-36-034-ORD-966 with the University of Maryland. This paper was presented to the American Mathematical Society, Feb. 28, 1953.

[2] Numbers in brackets, [ ], refer to the bibliography at the end of the paper.

*$u(x, y)$ uniformly on every closed arc of $S$ containing no point $Q$, then $u_A(x, y)$ approaches $u(x, y)$ uniformly on every closed point set of $\Omega$ containing no point $Q$.*

Both parts of Theorem 2.1 are conveniently proved by use of

LEMMA 2.2. *Under the conditions of Theorem* 2.1, *let $v(x, y)$ be harmonic in $R$, $|v(x, y)|$ not greater than $M$ in $R$, and* $\limsup |v(x, y)| \leqq \epsilon$ *as $(x, y)$ in $R$ approaches an arbitrary point of $S'$, where $S'$ is a closed subset of $S$ containing no point $Q$. Let $S_1$ be a closed subset of $S$ containing in its interior all the points of $S - S'$. Then on any closed subset $S_2$ of $\Omega - S_1$ we have*

$$|v(x, y)| \leqq M\omega(x, y) + \epsilon,$$

*where $\omega(x, y)$ is the harmonic measure at $(x, y)$ of $S_1$ with respect to $\Omega$.*

Lemma 2.2 is proved by merely studying lim sup and lim inf of $v(x, y)$ as $(x, y)$ in $R$ approaches points of $S$. The existence of $u(x, y)$ in Theorem 2.1 is shown by the Cauchy condition for uniform convergence from Lemma 2.2 by use of a bounded sequence of suitably chosen boundary values $F_k(x, y)$, each continuous on $S$, equal to $F(x, y)$ on $S$ except in a neighborhood of each point $Q$, where the neighborhoods shrink indefinitely. Of course if $S_2$ is fixed, $\omega(x, y)$ approaches zero uniformly on $S_2$ as the set $S_1$ shrinks and approaches the point set $Q$. The uniqueness of $u(x, y)$ is a consequence of Lemma 2.2. The latter part of Theorem 2.1 then follows readily from Lemma 2.2. For a discussion both of the Dirichlet problem and harmonic measure the reader may refer to Nevanlinna [8], Chapter II.

Henceforth throughout this paper we shall consider $R$ as the square $0 < x < 1$, $0 < y < 1$, and we shall assume that $F(x, y)$ satisfies the conditions given in Theorem 2.1 and that $F(x, y)$ tends to a limit as $(x, y)$, on $S$, approaches a point of discontinuity $Q$ in each of the two directions. At such a point $Q$ we ordinarily take the average of the two limits as the value of $F(x, y)$.

The most common finite difference analogue of the Dirichlet problem is constructed as follows. With *mesh size* $h = 1/A$, where $A$ is a positive integer, we seek a function $u_A(x, y)$ which satisfies the difference equation

$$\begin{aligned} &u_A(x + h, y) + u_A(x - h, y) + u_A(x, y + h) \\ &\qquad + u_A(x, y - h) - 4u_A(x, y) = 0 \end{aligned} \tag{2.1}$$

at the interior *nodes* $(x, y)$, that is points of $R$ such that $x$ and $y$ are integral multiples of $h$, and which equals $F(x, y)$ at the boundary nodes. Functions satisfying (2.1) are said to be *discrete harmonic* and are discussed in [3].

By linearity of the Dirichlet problem and its difference analogue it is sufficient for a study of the degree of convergence of $u_A(x, y)$ to $u(x, y)$ to assume that the boundary values are determined by $F_1(x, y)$, where $F_1(x, y)$ vanishes except when $y = 0$, and

$$\begin{aligned} &F_1(x, 0) = F(x, 0), && 0 < x < 1 \\ &F_1(0, 0) = \tfrac{1}{2} \operatorname{Lim}_{x \to 0+} F(x, 0), \qquad F_1(1, 0) = \tfrac{1}{2} \operatorname{Lim}_{x \to 1-} F(x, 0). \end{aligned} \tag{2.2}$$

Evidently $F(x, y)$ is the sum of $F_1(x, y)$ and three other similar functions.

Since the values of $u_A(x, y)$ at the corners of the square do not appear in

(2.1), we can double $F_1(x, y)$ at the points (0, 0) and (1, 0) without changing $u_A(x, y)$ or $u(x, y)$ except at these two points. With $F_1(x, y)$ thus modified we let, for convenience,

$$f(x) = F_1(x, 0), \qquad 0 \leqq x \leqq 1.$$

Evidently $f(x)$ is piecewise continuous on $I: 0 \leqq x \leqq 1$, that is to say, continuous on $I$ except possibly for a finite number of finite jumps. Moreover $f(x)$ satisfies the conditions

$$\begin{aligned} f(x) &= \tfrac{1}{2}[f(x+) + f(x-)], && 0 < x < 1 \\ f(0) &= f(0+), \qquad f(1) = f(1-) \end{aligned} \tag{2.3}$$

where $f(x+) = \operatorname{Lim}_{t \to x+} f(t)$ and $f(x-) = \operatorname{Lim}_{t \to x-} f(t)$.

Since our study of $u(x, y)$ and $u_A(x, y)$ will involve only subregions of the square not containing either of the points (0, 0) and (0, 1), it is sufficient for us to assume that on $S$

$$\begin{aligned} u(x, y) &= 0, && y > 0 \\ u(x, 0) &= f(x), && 0 \leqq x \leqq 1 \end{aligned}$$

and that for nodes on $S$

$$\begin{aligned} u_A(x, y) &= 0, && y > 0 \\ u_A(x, 0) &= f(x). \end{aligned}$$

We now prove

THEOREM 2.3. *The unique solution of the above Dirichlet problem is given by*

$$\begin{aligned} u(x, y) &= \textstyle\sum_{n=1}^{\infty} a_n \sin n\pi x \, W(n, y), && \text{if } y > 0, \\ u(x, y) &= f(x), && \text{if } y = 0, \end{aligned} \tag{2.4}$$

*where*

$$W(n, y) = \frac{\sinh n\pi(1 - y)}{\sinh n\pi}, \tag{2.5}$$

$$a_n = 2 \int_0^1 f(x) \sin n\pi x \, dx. \tag{2.6}$$

PROOF. The series in the right member of (2.4) converges for $y > 0$; let $u^*(x, y)$ denote the right member of (2.4). Let the functions $u_k(x, y)$, $k = 1, 2, \cdots$, be the arithmetic means of the first $k$ partial sums of (2.4). Since $(C, 1)$ summability is regular it follows that $u_k(x, y)$ approaches $u^*(x, y)$ for $y > 0$. Moreover, in any closed subinterval of $I$ not containing points $Q$, which are either points of discontinuity of $f(x)$ or the points $x = 0$, $x = 1$, $u_k(x, 0)$ approaches $f(x)$ uniformly. Therefore in any closed subset of $\Omega$ not containing any point $Q$ the sequence $u_k(x, y)$ approaches $u^*(x, y)$.

Moreover, for each $k$, $u_k(x, y)$ is harmonic in $R$, continuous in $\Omega$ and, since

by a well known property of $(C, 1)$ summability of Fourier series, $| u_k(x, 0) | \leqq M$, where $M = \text{L.U.B.} | f(x) |$, for $x \varepsilon I$, we have $| u_k(x, y) | \leqq M$ in $\Omega$. Therefore the hypotheses of the second part of Theorem 2.1 are satisfied and $u_k(x, y)$ converges to $u(x, y)$ uniformly in any closed subset of $\Omega$ containing no point $Q$. Therefore except at the points $Q$ we have $u(x, y) = u^*(x, y)$. But by (2.4) the equality also holds at the points $Q$, and the theorem is proved.

Theorem 2.3 is valid under the hypothesis of Theorem 2.1 on $F(x, y)$, without assuming $f(x+)$ and $f(x-)$ to exist.

For the difference equation problem it is known, see for instance [7], [9] or [12], that the unique solution is

$$u_A(x, y) = \sum_{n=1}^{A-1} b_n(A) \sin n\pi x \, W(m, y) \tag{2.7}$$

for nodes $(x, y)$ such that $y > 0$, where

$$b_n(A) = (2/A) \sum_{j=1}^{A} f(j/A) \sin (n\pi j/A), \tag{2.8}$$

and for each integer $n$ such that $1 \leqq n \leqq A - 1$ the number $m$ is defined implicitly by

$$\sinh m\pi/2A = \sin n\pi/2A. \tag{2.9}$$

To study the degree of convergence of $u_A(x, y)$ to $u(x, y)$ as $A \to \infty$ we define for each $y_0$ such that $1 \geqq y_0 > 0$

$$E(A, y_0) = \text{L.U.B.} | u_A(x, y) - u(x, y) |, \tag{2.10}$$

the least upper bound being taken over all mesh points $(x, y)$ such that $1 \geqq y \geqq y_0$, $0 \leqq x \leqq 1$. Typical results will be of the form

$$E(A, y_0) = O(\phi(A))$$

where $\phi(A)$ is to be evaluated as some non-negative function of $A$ which tends monotonically to zero as $A \to \infty$. The constant factor implied by $O(\phi(A))$ depends on $y_0$ but not on $A$.

**3. Preliminary results.** We now show that the degree of convergence of $u_A(x, y)$ to $u(x, y)$ depends, essentially, on the degree of convergence of the $b_n(A)$ in (2.8) to the Fourier coefficients $a_n$.

THEOREM 3.1. *If $f(x)$ is piecewise continuous and if there exist functions $\phi(A)$ and $\psi(n)$ of the positive integers $A$ and $n$, respectively, such that $\text{Lim}_{A\to\infty}\phi(A) = 0$, and such that for some $s > 0$, we have $\psi(n) = O(n^s)$, and*

$$| a_n - b_n(A) | \leqq \psi(n)\phi(A), \qquad (n = 1, 2, \cdots, A - 1), \tag{3.1}$$

*then*

$$E(A, y_0) = O(A^{-2}), \text{ if } \phi(A) = O(A^{-2})$$
$$= O(\phi(A)) \text{ otherwise.} \tag{3.2}$$

*Here $E(A, y_0)$ is defined by (2.10) for $1 \geqq y_0 > 0$.*

PROOF. Subtracting (2.4) from (2.7) we get for $y > 0$

$$| u(x, y) - u_A(x, y) | \leq E_1(x, y) + E_2(x, y) + E_3(x, y)$$

where

$$E_1(x, y) = | \sum_{n=A}^{\infty} a_n \sin n\pi x \, W(n, y) |$$

$$E_2(x, y) = | \sum_{n=1}^{A-1} b_n(A) \sin n\pi x \, (W(n, y) - W(m, y)) |$$

$$E_3(x, y) = | \sum_{n=1}^{A-1} (a_n - b_n(A)) \sin n\pi x \, W(n, y) | .$$

It is easy to show that

$$W(n, y) \leq e^{-n\pi y}, \qquad 0 \leq y \leq 1, \tag{3.3}$$

see for instance [9]. Next, since $f(x)$ is piecewise continuous, it follows that for some $M$ we have $| f(x) | \leq M$; hence $| a_n | \leq 2M$. Therefore

$$E_1(x, y) \leq 2M \sum_{n=A}^{\infty} e^{-n\pi y} \leq \frac{2Me^{-A\pi y_0}}{1 - e^{-\pi y_0}}, \qquad (y \geq y_0 > 0).$$

For fixed $y_0 > 0$, we have $E_1(x, y) = O(A^{-2})$.

Again, since $| f(x) | \leq M$, we have, by (2.8), $| b_n(A) | \leq 2M$. Moreover, in [12], Section 6, it is shown that

$$| W(m, y) - W(n, y) | \leq QA^{-2} y n^3 e^{-Bny}$$

where

$$Q = (\pi^3/12) \coth B, \qquad B = 2 \sinh^{-1}(1).$$

Therefore we have for $1 \geq y \geq y_0 > 0$,

$$E_2(x, y) \leq 2MQA^{-2} y \sum_{n=1}^{A-1} n^3 e^{-Bny}$$

$$\leq 2MQA^{-2} \sum_{n=1}^{\infty} n^3 e^{-By_0 n} \leq Q_2(y_0) A^{-2},$$

where $Q_2(y_0)$ depends only on $y_0$. This follows since the infinite series converges.

Finally, we have by hypothesis and (3.3)

$$E_3(x, y) \leq \phi(A) \sum_{n=1}^{A-1} \psi(n) e^{-n\pi y}$$

$$\leq \phi(A) \sum_{n=1}^{\infty} \psi(n) e^{-n\pi y_0}$$

for $y \geq y_0 > 0$. Since $\psi(n) = O(n^s)$ by hypothesis, it follows that the infinite series converges. The theorem now follows.

Next, we prove

THEOREM 3.2. *If $f(x)$ has the property that for some $K$ and some $\gamma > 1$, $| a_n | \leq K/n^\gamma$, $n = 1, 2, \cdots$, then*

$$E(A, y_0) = O(A^{-\gamma}), \qquad \text{if } \gamma \leq 2$$

$$= O(A^{-2}), \qquad \text{if } \gamma > 2. \tag{3.4}$$

PROOF. It is easily verified that the Fourier series for $f(x)$ converges absolutely and uniformly for all $x \in I$ to $f(x)$, and that $f(x)$ is continuous.

Now, it is known, see Wasow [13], that if $f(x) \equiv \sum_{n=1}^{\infty} a_n \sin n\pi x$ where the $a_n$ are given by (2.6), then

$$a_n - b_n(A) = -\sum_{k=1}^{\infty} (a_{2kA+n} - a_{2kA-n}). \tag{3.5}$$

Hence, we have for all $n$ such that $1 \leqq n \leqq A - 1$

$$\begin{aligned} |a_n - b_n(A)| &= |\textstyle\sum_{k=1}^{\infty} a_{2kA+n} - a_{2kA-n}| \\ &\leqq K \textstyle\sum_{k=1}^{\infty} [(2kA + n)^{-\gamma} + (2kA - n)^{-\gamma}] \\ &\leqq (2K/A^{\gamma}) \textstyle\sum_{k=1}^{\infty} (2k - 1)^{-\gamma} = O(A^{-\gamma}) \end{aligned} \tag{3.6}$$

since the infinite series converges. The result now follows by Theorem 3.1.

Obviously more delicate conditions than those of Theorem 3.2 can be used, with correspondingly more delicate results.

We shall also make use of the following theorem.

THEOREM 3.3. *Let $f(x)$ have the property that there exists a monotone decreasing sequence of positive numbers $\epsilon_1, \epsilon_2, \cdots$ tending to zero, and for each $N$ there exists a trigonometric sum $S_N(x)$ of the form*

$$S_N(x) = \sum_{n=1}^{N} \alpha_{N,n} \sin n\pi x$$

*such that*

$$|f(x) - S_N(x)| \leqq \epsilon_N$$

*uniformly for $0 \leqq x \leqq 1$. Then*

$$\begin{aligned} E(A, y_0) &= O(\epsilon_A) && \text{if } A^{-2} = O(\epsilon_A) \\ &= O(A^{-2}) && \text{if } \epsilon_A = O(A^{-2}). \end{aligned} \tag{3.7}$$

PROOF. By (2.6) and (2.8) we have as in [11], Sec. 5,

$$a_n - b_n(A) = 2\int_0^1 S_A(x) \sin n\pi x\, dx + 2\int_0^1 [f(x) - S_A(x)] \sin n\pi x\, dx$$

$$- (2/A) \sum_{j=1}^{A-1} S_A(j/A) \sin n\pi j/A - (2/A) \sum_{j=1}^{A-1} [f(j/A) - S_A(j/A)] \sin n\pi j/A.$$

But it is easy to show, (see for instance [4], Chapter IV), that for $n \leqq A - 1$, we have

$$(2/A) \sum_{j=1}^{A-1} S_A(j/A) \sin (n\pi j/A) = 2\int_0^1 S_A(x) \sin n\pi x\, dx = \alpha_{A,n}.$$

Therefore for $n \leqq A - 1$ we have

$$|a_n - b_n(A)| \leqq 4\epsilon_A. \tag{3.8}$$

The result follows from Theorem 3.1.

**4. Step functions and polynomials.** We shall now discuss the convergence of $u_A(x, y)$ to $u(x, y)$ as $A$ takes on a sequence of different values $A_1, A_2, \cdots$ such

that for each $k$, $A_{k+1}$ is an integral multiple of $A_k$. It is evident that $A_k$ is a multiple of $A_j$ for all $j \leq k$. Therefore, if for some $k$, the point $(x, y)$ is a node of the net with mesh size $A_k^{-1}$, then it will be a node of all finer nets of the sequence. Thus with a given sequence $\{A_k\}$ the points of $\Omega$ are divided into two classes: those which are nodes of some net and those which are not. We shall refer to points of the former class as *mesh points*.

Let us now consider the case where $f(x)$ is the characteristic function of a subinterval $[\alpha, \beta]$ of $I$.

THEOREM 4.1. *Let* $f(x) = 1$, $\alpha < x < \beta$, $f(\alpha) = f(\beta) = \frac{1}{2}$ *except that* $f(\alpha) = f(0+)$ *if* $\alpha = 0$, $f(\beta) = f(1-)$ *if* $\beta = 1$, *and* $f(x) = 0$ *elsewhere in the interval* $I$. *Then*

$$E(A, y_0) = O(A^{-1}). \tag{4.1}$$

*If both of the points* $(\alpha, 0)$, $(\beta, 0)$ *are mesh points then*

$$E(A, y_0) = O(A^{-2}). \tag{4.2}$$

PROOF. By Theorem 3.1 it is sufficient to study the difference $a_n - b_n(A)$ for each $n$. Let $p$, $q$ denote respectively the smallest integer not less than $A\alpha$ and the largest integer not greater than $A\beta$. By (2.6) we have at once

$$a_n = (2/n\pi)[\cos n\pi\alpha - \cos n\pi\beta]. \tag{4.3}$$

Next, it is easily seen that

$$\begin{aligned} \tfrac{1}{2}A\, b_n(A) &= \textstyle\sum_{j=1}^{A} f(j/A) \sin n\pi j/A \\ &= \textstyle\sum_{j=p}^{q} \sin n\pi j/A - [(1 - f(ph)) \sin n\pi ph + (1 - f(qh)) \\ &\qquad \sin n\pi qh] \end{aligned} \tag{4.4}$$

But by a well-known formula for the sum of a series of sines, see for instance [11], we have

$$\begin{aligned} \textstyle\sum_{j=p}^{q} \sin n\pi j/A &= \tfrac{1}{2} \cot (n\pi/2A) \cdot (\cos pn\pi/A - \cos qn\pi/A) \\ &\quad + \tfrac{1}{2}(\sin pn\pi/A + \sin qn\pi/A). \end{aligned} \tag{4.5}$$

Now if the points $(\alpha, 0)$ and $(\beta, 0)$ are mesh points and if $A$ is sufficiently large so that they are nodes of the net with mesh size $1/A$, then the second term of the last member of (4.4) equals $-\frac{1}{2}[\sin n\pi p/A + \sin n\pi q/A]$, and we have

$$b_n(A) = (1/A) \cot (n\pi/2A)(\cos n\pi\alpha - \cos n\pi\beta). \tag{4.6}$$

We now observe that for $0 < x \leq \frac{1}{2}\pi$ we have

$$0 \leq (1/x) - \cot x \leq (4/\pi^2)\, x. \tag{4.7}$$

For the series for $\cot x$

$$\cot x = x^{-1} - \tfrac{1}{3}x - \tfrac{1}{45}x^3 - \cdots$$

has radius of convergence $\pi$, see for instance [6], page 418. Therefore the series for $P(x)$ where $P(x)$ is defined by

$$\cot x = x^{-1} - xP(x)$$

has the same radius of convergence, and moreover, all coefficients in the expansion of $P(x)$ are positive, ([6], page 204). Therefore $P(x)$ is an increasing function of $x$, and its maximum value in the interval $0 \leqq x \leqq \pi/2$ is given by $(2/\pi)^2$. Therefore (4.7) follows. But by (4.3) and (4.6) we have

$$a_n - b_n(A) = (\cos n\pi\alpha - \cos n\pi\beta)[(2/n\pi) - A^{-1}\cot(n\pi/2A)]. \tag{4.8}$$

By (4.7) we have

$$0 \leqq (2/n\pi) - A^{-1}\cot(n\pi/2A) \leqq 2n/\pi A^2, \tag{4.9}$$

and

$$|a_n - b_n(A)| \leqq |\cos n\pi\alpha - \cos n\pi\beta|\,(2n/\pi A^2). \tag{4.10}$$

Thus (4.2) follows from Theorem 3.1.

To prove (4.1) we rewrite (4.5) in the form

$$\sum_{j=p}^{q} \sin\frac{n\pi j}{A} = \tfrac{1}{2}\csc\frac{n\pi}{2A}\left(\cos\frac{(p-\tfrac{1}{2})n\pi}{A} - \cos\frac{(q+\tfrac{1}{2})n\pi}{A}\right).$$

By (4.4) we have

$$b_n(A) = \frac{1}{A}\csc\frac{n\pi}{2A}\left(\cos\frac{(p-\tfrac{1}{2})n\pi}{A} - \cos\frac{(q+\tfrac{1}{2})n\pi}{A}\right) + \theta$$

$$= \frac{1}{A}\csc\frac{n\pi}{2A}\left[\cos\left(\alpha n\pi + \frac{\epsilon_1 n\pi}{A}\right) - \cos\left(\beta n\pi + \frac{\epsilon_2 n\pi}{A}\right)\right] + \theta$$

where $(A/2)\theta = -[(1 - f(ph))\sin n\pi ph + (1 - f(qh))\sin n\pi qh]$ and where $\epsilon_1 = (p - \tfrac{1}{2}) - A\alpha$, $\epsilon_2 = (q - \tfrac{1}{2}) - A\beta$. Evidently $|\epsilon_1| < 1$, $|\epsilon_2| < 1$, and $|\theta| \leqq 2/A$. Actually $\theta$ vanishes if neither of the points $(\alpha, 0)$, $(\beta, 0)$ are nodes.

Now using the power series expansion for the function $\csc x$, it can be shown that

$$0 \leqq \csc x - (x^{-1}) \leqq Hx, \qquad 0 < x \leqq \pi/2,$$

where $H = (2\pi - 4)/\pi^2$. Hence for $n \leqq A$ we have

$$0 \leqq \csc(n\pi/2A) - (2A/n\pi) \leqq H(n\pi/2A).$$

Moreover

$$|\cos[n\pi\alpha + (\epsilon_1 n\pi/A)] - \cos n\pi\alpha| \leqq |\epsilon_1|\, n\pi/A,$$

$$|\cos[n\pi\beta - (\epsilon_2 n\pi/A)] - \cos n\pi\beta| \leqq |\epsilon_2|\, n\pi/A.$$

It then follows that for all $n \leqq A$

$$|a_n - b_n(A)| \leqq Hn\pi/A^2 + (\pi/A)[|\epsilon_1| + |\epsilon_2|] + (2/A) \leqq Q_1/A,$$

where $Q_1$ is independent of $A$ and $n$. Equation (4.1) follows by Theorem 3.1.

By linearity of the difference and differential equations we have

THEOREM 4.2. *Let $f(x)$ be constant in each of the intervals $0 \leqq x < x_1$, $x_1 < x <$*

*$x_2, \cdots, x_{n-1} < x < x_n, x_n < x \leqq 1$, and for each $i$ let $f(x_i) = \frac{1}{2}[f(x_i+) + f(x_i-)]$. Then*

$$E(A, y_0) = O(A^{-1}) \tag{4.11}$$

*and if all the points $x_1, x_2, \cdots, x_n$ are mesh points*

$$E(A, y_0) = O(A^{-2}). \tag{4.12}$$

Next let us consider the case where $f(x)$ is a linear function, say, $cx + d$ where $c$, $d$ are constants. By Theorem 4.1, letting $\alpha = 0$, $\beta = 1$, it is sufficient to assume $d = 0$, for otherwise we could subtract $d$ from $f(x)$ and the degree of convergence with $f(x) \equiv d$ is $O(A^{-2})$.

THEOREM 4.3. *If $f(x) \equiv cx$ then*

$$E(A, y_0) = O(A^{-2}). \tag{4.13}$$

PROOF. From (2.6) we get $a_n = (-1)^{n+1}\, 2c/n\pi$ and by (2.8)

$$b_n(A) = (2c/A^2) \sum_{j=1}^{A-1} j \sin n\pi j/A.$$

The last expression can be summed by differentiating with respect to $\lambda$ the identity

$$-\sum_{j=1}^{A-1} \cos \lambda j = -\frac{\sin \frac{1}{2}(A-1)\lambda \cos \frac{1}{2}A\lambda}{\sin \frac{1}{2}\lambda}$$

thereby obtaining, after reduction, and letting $\lambda = n\pi/A$

$$b_n(A) = (-1)^{n+1}\, (c/A) \cot (n\pi/2A).$$

Using (4.9) one can show that

$$|a_n - b_n(A)| \leqq (2cn/\pi A^2), \tag{4.14}$$

and the result follows by Theorem 3.1.

We remark that for Theorem 4.3, it is not required that $A$ tend to infinity in the manner prescribed at the beginning of this section. This is true for polynomials as well. In fact we have

THEOREM 4.4. *If $f(x) \equiv x^p$, $(p \geqq 2)$, then*

$$E(A, y_0) = O(A^{-2}). \tag{4.15}$$

Let $f_2(x) \equiv x^p - x$. Evidently $f_2(0) = f_2(1) = 0$ and $f_2(x)$ has a bounded second derivative in the interior of $I$. It is easy to show by repeated integration by parts that the Fourier coefficients tend to zero like $n^{-2}$ as $n \to \infty$; hence the theorem follows immediately by Theorems 3.2 and 4.3. The extension to the case where $f(x)$ is an arbitrary polynomial is immediate.

**5. Piecewise continuous functions.** For the case of continuous boundary values we prove

THEOREM 5.1. *Let $f(x)$ be continuous in the interval $0 \leqq x \leqq 1$. If $f$ is not constant, then*

$$E(A, y_0) = O(\omega(A^{-1})) \tag{5.1.}$$

*where $\omega(\delta)$ is a modulus of continuity of $f(x)$, in the sense that for points of $I$ the inequality $|x_1 - x_2| \leqq \delta$ implies $|f(x_1) - f(x_2)| \leqq \omega(\delta)$.*

PROOF. As before we let $g(x) \equiv f(x) \sin n\pi x$. Now, if $|x_1 - x_2| < \delta$, then

$$|g(x_2) - g(x_1)| \leqq |f(x_1) - f(x_2)|\,|\sin n\pi x_1| + |f(x_2)|\,|\sin n\pi x_2 - \sin n\pi x_1|$$
$$\leqq \omega(\delta) + Mn\pi\delta$$

since $|\sin n\pi x_2 - \sin n\pi x_1| \leqq n\pi\delta$. Here $M$ denotes the maximum of $|f(x)|$. Therefore, $w_g(\delta)$, a modulus of continuity of $g(x)$, is

$$\omega_g(\delta) \equiv \omega(\delta) + Mn\pi\delta.$$

Now by Theorem 1 of [11] we have

$$\tfrac{1}{2}|a_n - b_n(A)| = \left|\int_0^1 g(x)\,dx - (1/A)\sum_{j=1}^{A} g(j/A)\right| \leqq \omega_g(A^{-1}) \tag{5.2}$$
$$\leqq \omega(A^{-1}) + Mn\pi A^{-1}.$$

Since $f(x)$ is not constant, $\omega(A^{-1}) \neq 0$ and $A^{-1} = O(\omega(A^{-1}))$, see Theorem 2.4 of [1]. We have

$$\tfrac{1}{2}|a_n - b_n(A)| \leqq Mn\pi\omega(A^{-1})\left[1 + \frac{A^{-1}}{\omega(A^{-1})}\right].$$

The theorem follows from Theorem 3.1.

COROLLARY. *If $f(x)$ satisfies a Lipschitz condition in $I$ of order $\alpha$ for some $\alpha$ such that $0 < \alpha \leqq 1$, then*

$$E(A, y_0) = O(A^{-\alpha}). \tag{5.3}$$

Next, we prove

THEOREM 5.2. *Let $f(x)$ be continuous in $I$ and have a derivative $f'(x)$ except perhaps for a finite number of points of $I$ and let $f'(x)$ be piecewise continuous in $I$. Let $h_1(x)$ be the unique function such that $h_1(0) = 0$, $h_1(x)$ is continuous in $I$, and the function $h_2(x) \equiv f'(x) - h_1(x)$ is a step function. If $f(0) = f(1) = 0$ and if $\omega_1(\delta)$ is a modulus of continuity of $h_1(x)$ in $I$ then*

$$E(A, y_0) = O(A^{-1}\omega_1(2A^{-1})), \qquad \text{if } \omega_1(\delta) \neq 0,$$
$$= O(A^{-2}) \qquad \text{if } \omega_1(\delta) = 0. \tag{5.4}$$

PROOF. Evidently $f(x) \equiv f_1(x) + f_2(x)$ where $f_1(x) \equiv \int_0^x h_1(x)\,dx$ and $f_2(x) \equiv \int_0^x h_2(x)\,dx$. Moreover, for $0 < x < 1$, $f_1'(x) \equiv h_1(x)$ and $f_2'(x) \equiv h_2(x)$ except where $f'(x)$ is not defined or is not continuous.

Since $f_2(0) = 0$, the function $f_3(x) \equiv f_2(x) - f_2(1)x$ vanishes at the end points of $I$. This function is a broken line function and by the method used in [5], pp. 13–14, it is readily shown that its Fourier coefficients are of order $n^{-2}$ as $n \to \infty$. By Theorems 3.2 and 4.3 the theorem is true for $f_2(x)$.

Next, since $f_1(0) = 0$, the function $f_4(x) \equiv f_1(x) - f_1(1)x$ vanishes at the

end points of $I$. We can extend the domain of $f_4(x)$ so that $f_4(-x) \equiv -f_4(x)$, $f_4(x+2) \equiv f_4(x)$, and $f_4(x)$ is continuous everywhere. Since $f_1'(x)$, and hence $f_4'(x)$, are uniformly continuous in the interior of $I$, it follows that $f_4(x)$ has a right hand and a left hand derivative, respectively, for $x = 0$ and $x = 1$. Hence, by the construction of $f_4(x)$, $f_4'(x)$ exists for $x = 0$ and $x = 1$, and, in fact, $f_4'(x)$ exists and is continuous everywhere. Moreover $f_4'(x)$ has a modulus of continuity, $\omega_4(\delta)$, which does not exceed $2\omega_1(\delta)$, for $\delta \leqq 1$. By [4], p. 12, for each $k$, there exists a trigonometric (sine) sum $S_k(x)$ of order $k$ such that

$$|f_4(x) - S_k(x)| = O(k^{-1}\omega_4(2k^{-1})) \tag{5.5}$$

as $k \to \infty$. The theorem now follows by Theorems 3.3 and 4.3.

COROLLARY. *Under the hypotheses of Theorem* 5.2 *and the additional requirement that* $h_1(x)$ *satisfy a Lipschitz condition of order* $\alpha$, $0 < \alpha \leqq 1$, *in every interval of continuity, then*

$$E(A, y_0) = O(1/A^{1+\alpha}). \tag{5.6}$$

Next, we consider the case where $f'(x)$ has bounded variation.

THEOREM 5.3. *Let* $f(x)$ *be continuous in* $I$, *let* $f(0) = f(1) = 0$, *and let* $f'(x)$ *exist except for a finite number of points. If* $f'(x)$ *has bounded variation in* $I$, *then*

$$E(A, y_0) = O(1/A^2). \tag{5.7}$$

PROOF. Using the indicated generalization of Corollary I of [4] page 50, we have

$$|a_n| \leqq 2V/n^2\pi^2 \tag{5.8}$$

where $V$ is the total variation of $f'(x)$. The result now follows from Theorem 3.2.

Now, given a function $f(x)$ continuous except for a finite number of finite jumps we can express $f(x)$ in the form

$$f(x) \equiv f_1(x) + f_2(x) + f_3(x)$$

where $f_1(0) = f_1(1) = 0$ and $f_1(x)$ is continuous in the closed interval $0 \leqq x \leqq 1$, $f_2(x)$ is a step function, and $f_3(x)$ is a linear function. Using Theorems 4.2 and 4.3 we can assert the following

THEOREM 5.4. *Theorems* 5.2 *and* 5.3 *remain valid if we remove the requirement* $f(0) = f(1) = 0$. *Moreover, if* $f(x)$ *is continuous except for a finite number of finite jumps then*

(a) *Theorem* 5.1 *remains valid*

(b) *Theorems* 5.2 *and* 5.3 *remain valid if all jumps of* $f(x)$ *occur at mesh points*

(c) *Without restriction on the points at which* $f(x)$ *has finite jumps Theorems* 5.2 *and* 5.3 *remain valid if we replace the conclusion of each by*

$$E(A, y_0) = O(A^{-1}). \tag{5.9}$$

**6. Functions of bounded variation.** Our next result is

THEOREM 6.1. *If* $f(x)$ *is piecewise continuous and has bounded variation in the interval* $0 \leqq x \leqq 1$, *then*

$$E(A, y_0) = O(A^{-1}). \tag{6.1}$$

PROOF. Let $g(x) \equiv f(x) \sin n\pi x$. If $V$ is the total variation of $f(x)$ then the total variation of $g(x)$ does not exceed $V + 2Mn$. Using Theorem 2 of [11] we get

$$\frac{1}{2}\,|\,a_n - b_n(A)\,| = \left| \int_0^1 g(x)\,dx - \frac{1}{A}\sum_{j=1}^{A} g\left(\frac{j}{A}\right) \right| \leqq \frac{V + 2Mn}{A}. \tag{6.2}$$

The result now follows by Theorem 3.1.

We remark that Theorems 2.1 and 2.3 can be extended to include cases where $f(x)$ has an infinite number of jumps and has bounded variation. Theorem 6.1 is also true under these weaker hypotheses.

**7. Continuous functions; an example.** *Even for continuous boundary values, the degree of convergence may be arbitrarily slow*, as we now prove. We let $u(x, y)$ be the solution of the Dirichlet problem and denote the solution of the difference equation with mesh size $A_j^{-1} = 2^{-j}$ by $u^{(j)}(x, y)$.

THEOREM 7.1. *Let the sequence of positive numbers $\delta_j$ approach zero monotonically as $j$ becomes infinite. Then there exists some $f(x)$, continuous in $I$ with $f(0) = f(1) = 0$, such that $u(x, 0) \equiv f(x)$ and*

$$\overline{\operatorname{Lim}}_{j\to\infty}\, \Delta_j/\delta_j = \infty \tag{7.1}$$

*where*

$$\Delta_j = u^{(j)}(\tfrac{1}{2}, \tfrac{1}{2}) - u(\tfrac{1}{2}, \tfrac{1}{2}). \tag{7.2}$$

PROOF. For each positive integer $k$ we define a continuous function $u_k(x, 0)$ in $I$ equal to unity in the interior points of the grid with mesh size $A^{-k}$, satisfying the inequalities $0 \leqq u_k(x, 0) \leqq 1$ on $I$ and such that the corresponding solution $u_k(x, y)$ of the Dirichlet problem satisfies the condition $u_k(\frac{1}{2}, \frac{1}{2}) = \frac{1}{8}$.

We now prove the existence of $u_k(x, 0)$. With mesh size $A_k^{-1}$, define $v_k(x, 0)$ as unity in the interior nodes of $I$, as zero outside the $\delta$-neighborhoods of these points, and as linear in each interval in which the function is not otherwise defined. Then the corresponding harmonic function $v_k(x, y)$, using boundary values zero on the remaining three sides of $\Omega$, approaches zero in the point $(\frac{1}{2}, \frac{1}{2})$ as $\delta \to 0$. Choose $\delta$ so small that for the corresponding harmonic function $v_k^{(1)}(x, y)$ we have $v_k^{(1)}(\frac{1}{2}, \frac{1}{2}) < \frac{1}{8}$. With mesh size $A_k^{-1}$ define $V_k(x, 0)$ as unity except in $\delta$-neighborhoods of the midpoints $B$ of the sub-intervals of $I$ bounded by successive interior nodes, and in the first and last intervals bounded by successive nodes define $V_k(x, 0)$ as unity except in one-sided $\delta$-neighborhoods of $x = 0$ and $x = 1$ respectively; define $V_k(x, 0)$ as zero in the points 0, 1 and $B$, and as linear in each interval in which the function is not otherwise defined. Then the corresponding harmonic function $V_k(x, y)$, using boundary values zero on the remaining sides of the square, approaches $\frac{1}{4}$ in the point $(\frac{1}{2}, \frac{1}{2})$ as $\delta \to 0$. Choose $\delta$ so small that for the corresponding harmonic function $v_k^{(2)}(x, y)$ we have $v_k^{(2)}(\frac{1}{2}, \frac{1}{2}) > \frac{1}{8}$; we need merely set

$$u_k(x, 0) = \frac{[8v_k^{(2)}(\frac{1}{2}, \frac{1}{2}) - 1]v_k^{(1)}(x, 0) + [1 - 8v_k^{(1)}(\frac{1}{2}, \frac{1}{2})]v_k^{(2)}(x, 0)}{8[v_k^{(2)}(\frac{1}{2}, \frac{1}{2}) - v_k^{(1)}(\frac{1}{2}, \frac{1}{2})]}.$$

Evidently, we have $0 \leqq u_k(x, 0) \leqq 1$ for $0 \leqq x \leqq 1$.

In order to assure that the boundary values are continuous at every boundary point of the square including the corners, we require $u_k(x, 0) = u_k(x, 1) = 0$. We note that in the above construction $u_k(x, 0)$ fulfills this condition.

The corresponding discrete harmonic function $u_k^{(j)}(x, y)$ for mesh size $A_j^{-1}$ satisfies

$$\begin{aligned} u_k^{(j)}(\tfrac{1}{2}, \tfrac{1}{2}) &= \tfrac{1}{4} && j \leq k \\ 0 < u_k^{(j)}(\tfrac{1}{2}, \tfrac{1}{2}) &\leq \tfrac{1}{4} && j > k. \end{aligned} \tag{7.3}$$

These inequalities follow from the symmetry of the problem with respect to the four sides of the square. We set $a_k^{(j)} = u_k^{(j)}(\frac{1}{2}, \frac{1}{2})$, so it follows from the continuity of $u_k(x, 0)$ that for each $k$ we have $a_k^{(j)} \to u_k(\frac{1}{2}, \frac{1}{2}) = \frac{1}{8}$ as $j \to \infty$. This follows by Theorem 5.1.

We now introduce the function

$$u(x, 0) = \sum_{k=1}^{\infty} 2^{-k} u_{n_k}(x, 0)$$

where the positive integers $n_k$ become infinite monotonically, and are to be specified later. By the Weierstrass $M$-test, this series converges uniformly on $I$, so $u(x, 0)$ is continuous on $I$. Likewise the series

$$u(x, y) = \sum_{k=1}^{\infty} 2^{-k} u_{n_k}(x, y)$$

converges uniformly in the closure of the square to a function harmonic in the interior of the square. (See Theorem 2.1.)

We have at once from (7.3)

$$\begin{aligned} \Delta_j = u^{(j)}(\tfrac{1}{2}, \tfrac{1}{2}) - u(\tfrac{1}{2}, \tfrac{1}{2}) &= \sum_{k=1}^{\infty} 2^{-k} a_{n_k}^{(j)} - \tfrac{1}{8} \sum_{k=1}^{\infty} 2^{-k} \\ &= \sum_{n_k < j} (a_{n_k}^{(j)} - \tfrac{1}{8}) 2^{-k} + \tfrac{1}{8} \sum_{n_k \geq j} 2^{-k}, \end{aligned}$$

which with the choice $j = n_\nu$ becomes

$$\Delta_{n_\nu} = \sum_{k<\nu} (a_{n_k}^{(n_\nu)} - \tfrac{1}{8}) 2^{-k} + \tfrac{1}{8} \sum_{k \geq \nu} 2^{-k}.$$

For definiteness we choose $n_1 = 1$ and then for $\nu > 1$ choose the successive $n_\nu$ $(> n_{\nu-1})$ so large that we have

$$\left| \sum_{k=1}^{\nu-1} (a_{n_k}^{(n_\nu)} - \tfrac{1}{8}) 2^{-k} \right| < \tfrac{1}{16} 2^{1-\nu},$$

which is possible because $a_{n_k}^{(j)} \to \frac{1}{8}$ as $j \to \infty$, and choose $n_\nu$ also so large that

$$\delta_{n_\nu} < 1/16\nu 2^{\nu-1}$$

which is possible since $\delta_j \to 0$. Then we have

$$\Delta_{n_\nu} > \tfrac{1}{16} 2^{1-\nu} > \nu \delta_{n_\nu},$$

hence

$$\Delta_{n_\nu} / \delta_{n_\nu} \to \infty$$

and the theorem is established.

The theorem involves merely degree of convergence at a particular point,

namely $(\frac{1}{2}, \frac{1}{2})$, but thereby includes results on degree of uniform convergence in any region containing that point. Of course an arbitrary mesh point interior to the given square can be treated by similar methods.

## BIBLIOGRAPHY


1. H. M. Elliott, *On approximation to functions satisfying a generalized continuity condition*, Trans. Amer. Math. Soc., vol. 71, pp. 1–23, 1951.
2. S. Gerschgorin, *Fehlerabschätzung für das Differenzenverfahren zur Lösung partieller Differentialgleichungen*, Zeits. f. angew. Math. u. Mech., vol. 10, pp. 373–382, 1930.
3. H. Heilbronn, *On discrete harmonic functions*, Proc. Camb. Phil. Soc., vol. 45, pp. 194–206, 1949.
4. D. Jackson, *The theory of approximation*, Amer. Math. Soc. Coll. Pub., vol. XI, 1930.
5. D. Jackson, *Fourier series and orthogonal polynomials*, Mathematical Association of America Carus Monograph No. 6, 1941.
6. K. Knopp, *Theory and application of infinite series*, translation from the Second German Edition by R. C. H. Young, Hafner Pub. Co.
7. J. Le Roux, *Sur le problème de Dirichlet*, Jour. de Math., vol. 10, pp. 189–230, 1914.
8. R. Nevanlinna, *Eindeutige analytische Funktionen*, Springer, Berlin, 1936.
9. H. Phillips and N. Wiener, *Nets and the Dirichlet problem*, J. Math. and Phys., vol. 2, pp. 105–124, 1923.
10. P. C. Rosenbloom, *On the difference equation method for solving the Dirichlet problem*, paper No. 25 of the Proceedings of a Symposium on the Construction and Applications of Conformal Maps, Los Angeles, 1949, (National Bureau of Standards Applied Mathematics Series 18).
11. J. L. Walsh and W. E. Sewell, *Note on degree of approximation to an integral by Riemann sums*, Amer. Math. Monthly, vol. 44, pp. 155–160, 1937.
12. J. L. Walsh and David Young, *On the accuracy of the numerical solution of the Dirichlet problem by finite differences*, Journal of Research of the National Bureau of Standards, vol. 51, pp. 343–363, 1953.
13. W. Wasow, *On the truncation error in the solution of Laplace's equation by finite differences*, Journal of Research of the National Bureau of Standards, vol. 48, pp. 345–348, 1952.


Harvard University
University of Maryland

(Received May 18, 1953)

# THE LOADED VIBRATING NET AND RESULTING BOUNDARY-VALUE PROBLEMS FOR A PARTIAL DIFFERENCE EQUATION OF THE SECOND ORDER

By Tomlinson Fort

**1. The Equation of Motion.** Let us be given a rectangular net of weightless cords loaded at each point of intersection $(x_i, y_j)$ with a particle of mass $m$. We shall furthermore assume vibrations small and perpendicular to the plane of the net when the net is at rest. We thus assume the tension in each string constant throughout its length. A portion of the net is shown in the accompanying figure. Denote the tension in the string over $(x_i, y_{j-1})$, $(x_i, y_j)$, $(x_i, y_{j+1})$ by $T_i$ and in the string over $(x_{i-1}, y_j)$ $(x_i, x_j)$, $(x_{i+1}, yj)$ by $\overline{T}_j$. We assume $T_i$ and $\overline{T}_j$ independent of $t$. We also assume vibrations so small that we can replace sin $\theta$ by tan $\theta$ where $\theta$ is an angle of the type indicated. Denote the vertical coordinate of $P_{ij}$ by $u_{ij}$. Under the circumstances, as outlined, the motion of $P_{ij}$ is determined by the following differential equation.

$$m\frac{d^2u_{ij}}{dt^2} = \overline{T}_j\frac{u_{i+1,j} - u_{ij}}{x_{i+1} - x_i} - \overline{T}_j\frac{u_{ij} - u_{i-1,j}}{x_i - x_{i-1}} + T_i\frac{u_{i,j+1} - u_{ij}}{y_{j+1} - y_j} - T_i\frac{u_{ij} - u_{i,j-1}}{y_j - y_{j-1}}, \qquad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n \tag{1}$$

For convenience in writing we let

$$b(i, j) = \frac{\overline{T}_j}{m(x_{i+1} - x_i)}, \qquad k(i, j) = \frac{T_i}{m(y_{j+1} - y_j)}$$

Equations (1) can then be written

$$\begin{aligned} d^2u_{ij}/dt^2 &= \Delta_i\{b(i-1, j)\Delta_i u_{i-1,j}\} \\ &\quad + \Delta_j\{k(i, j-1)\Delta_j u_{i,j-1}\} \quad i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, n. \end{aligned} \tag{2}$$

Equations (2) are a set of ordinary differential equations with constant coefficients. We proceed by letting $u_{ij} = e^{st}v(i, j)$. Substitute in (2) and divide by $e^{st}$. We obtain

$$\Delta_i\{b(i-1, j)\Delta_i v(i-1, j)\} + \Delta_j\{k(i, j-1)\Delta_j v(i, j-1)\} - s^2v(i, j) = 0 \qquad i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, n.$$

For future convenience we replace $-s^2$ by $\lambda$ and write

$$\begin{aligned} &\Delta_i\{b(i-1, j)\Delta_i v(i-1, j)\} \\ &\quad + \Delta_j\{k(i, j-1)\Delta_j v(i, j-1)\} + \lambda v(i, j) = 0 \end{aligned} \tag{3}$$

We note that $b(i, j) > 0$ and $k(i, j) > 0$. We have here $mn$ linear algebraic equations in $v(1, 1) \cdots v(m, n)$ which must be satisfied in order that (2) have

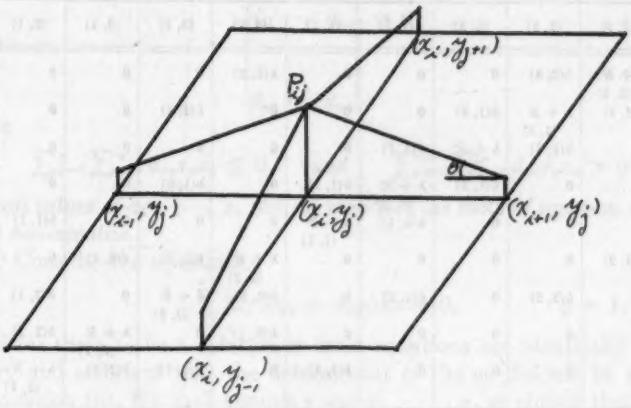



FIG. 1

a solution of the form indicated. We shall impose the additional restriction $u_{0j} = u_{i0} = u_{m+1,j} = u_{i,n+1} = 0$; $i = 1, \cdots, m$; $j = 1, \cdots, n$. These necessitate

$$v(0, j) = v(i, 0) = v(m + 1, j) = v(i, n + 1) = 0 \tag{4}$$

$$i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, n.$$

We designate by the term fundamental rectangle the set of points $i, j$ where $i = 0, \cdots, m + 1, j = 0, \cdots, n + 1$ with the exception of the point $(0, 0)$ $(0, n + 1)$ $(m + 1, 0)$ $(m + 1, n + 1)$

We shall study solutions of (3) subject to boundary conditions (4).

**2. The Characteristic Equation.** Write down equations (3) for $i = 1, \cdots, m$; $j = 1, 2, \cdots, n$ and then substitute in them form (4). There results a set of $mn$ homogeneous linear equations in $v(1, 1) \cdots v(m, n)$. In order for these to have a solution not identically zero it is necessary and sufficient that the determinant of the coefficients vanish. This determinant equated to zero is an equation in $\lambda$ to be called the characteristic equation for our problem. Its roots will be called characteristic values. The determinant itself will be called the characteristic determinant. The order in which these equations and the variable in them are arranged is not of great importance. However, the following order is adopted:

$$v(m, n), \cdots, v(1, n), v(1, n - 1), \cdots, v(1, 1), v(m, n - 1), \cdots,$$

$$v(2, n - 1), v(m, n - 2), \cdots, v(2, m - 2), \cdots, v(m, 1), \cdots v(2, 1)$$

Each equation will be designated by the pair of numbers, $(i, j)$ and the $v$'s will be arranged in the same order as the equations.

| | (3, 3) | (2, 3) | (1, 3) | (1, 2) | (1, 1) | (3, 2) | (2, 2) | (3, 1) | (2, 1) | |
|---|---|---|---|---|---|---|---|---|---|---|
| (3, 3) | $\lambda + R_{(3,3)}$ | $b(2,3)$ | 0 | 0 | 0 | $k(3,2)$ | 0 | 0 | 0 | |
| (2, 3) | $b(2,3)$ | $\lambda + R_{(2,3)}$ | $b(1,3)$ | 0 | 0 | 0 | $k(2,2)$ | 0 | 0 | |
| (1, 3) | 0 | $b(1,3)$ | $\lambda + R_{(1,3)}$ | $k(1,2)$ | 0 | 0 | 0 | 0 | 0 | |
| (1, 2) | 0 | 0 | $k(1,2)$ | $\lambda + R_{(1,2)}$ | $k(1,1)$ | 0 | $b(1,2)$ | 0 | 0 | |
| (1, 1) | 0 | 0 | 0 | $k(1,1)$ | $\lambda + R_{(1,1)}$ | 0 | 0 | 0 | $b(1,1)$ | $= 0.$ |
| (3, 2) | $k(3,2)$ | 0 | 0 | 0 | 0 | $\lambda + R_{(3,2)}$ | $b(2,2)$ | $k(3,1)$ | 0 | |
| (2, 2) | 0 | $k(2,2)$ | 0 | $b(1,2)$ | 0 | $b(3,2)$ | $\lambda + R_{(2,2)}$ | 0 | $k(2,1)$ | |
| (3, 1) | 0 | 0 | 0 | 0 | 0 | $k(3,1)$ | 0 | $\lambda + R_{(3,1)}$ | $b(2,1)$ | |
| (2, 1) | 0 | 0 | 0 | 0 | $b(1,1)$ | 0 | $k(2,1)$ | $b(2,1)$ | $\lambda + R_{(2,1)}$ | |

We write equation (3) in ordinary form. It is

$$b(i,j)v(i+1,j) + b(i-1,j)v(i-1,j) + k(i,j)v(i,j+1)$$

$$+ k(i,j-1)v(i,j-1) + [\lambda - \{b(i,j) + b(i-1,j) \tag{5}$$

$$+ k(i,j) + k(i,j-1\}]v(i,j) = 0; \quad b(i,j) > 0, \quad k(i,j) > 0.$$

For brevity we let

$$R(i,j) = -\{b(i,j) + b(i-1,j) + k(i,j) + k(i,j-1)\}$$

THEOREM 1. *The characteristic determinant is symmetrical with reference to its principal diagonal.*

*Proof*: For example by (5) we see that $b(i, j)$ occurs in the $(i, j)$ row and $(i + 1, j)$ column also in the $(i, j)$ column and $(i + 1, j)$ row.

We shall write down the characteristic equation for $m = n = 3$. This is for simplicity in exposition only. No arguments at any time will be special to this case.

We now prove a lemma following closely a proof in Perron.*

LEMMA: *Consider the equation.*

$$\begin{vmatrix} a_{11} + d_{11}z & \cdots & a_{1n} + d_{1n}z \\ \cdots\cdots\cdots & & \cdots\cdots\cdots \\ \cdots\cdots\cdots & & \cdots\cdots\cdots \\ a_{n1} + d_{n1}z & \cdots & a_{nn} + d_{nn}z \end{vmatrix} = 0 \tag{6}$$

Suppose that $a_{11}, \cdots, a_{nn}, \cdots, d_{11}, \cdots, d_{nn}$, are real and that $a_{pq} = a_{qp}$ *and* $d_{pq} = d_{qp}$ and that

* Algebra von Oskar Perron II s. 20.

$$\begin{vmatrix} d_{11} \cdots d_{1n} \\ \cdots\cdots\cdots \\ d_{n1} \cdots d_{nn} \end{vmatrix} \neq 0$$

and that

$$\sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}x_p x_q \leqq 0 \quad \text{and} \quad \sum_{p=1}^{n}\sum_{q=1}^{n} d_{pq}x_p x_q > 0$$

*for all real values of* $x_1, \cdots, x_n$ not all zero *then the roots of equation* (6) *are all real and non-negative.*

*Proof:* Consider the equations

$$\sum_{p=1}^{n} (a_{pq} + d_{pq}z)x_p = 0, \qquad q = 1, 2, \cdots, n \tag{7}$$

In order for there to be a solution of these equations not identically zero it is necessary and sufficient that the determinant of the coefficients be zero. This yields equation (6). We shall assume $z$ and $x_1, \cdots, x_n$ so chosen that $x_1, \cdots, x_n$ are not all zero and yet that equations (7) are satisfied. Suppose that $x_q$ is complex and denote its conjugate by $\bar{x}_q$. Then multiply equations (7) through by $\bar{x}_q$ and sum form 1 to $n$. We have

$$\sum_{q=1}^{n}\sum_{p=1}^{n} (a_{pq} + d_{pq}z)x_p\bar{x}_q = 0$$

or

$$z\sum_{q=1}^{n}\sum_{p=1}^{n} d_{pq}x_p\bar{x}_q = -\sum_{q=1}^{n}\sum_{p=1}^{n} a_{pq}x_p\bar{x}_q$$

Moreover since $a_{pq} = a_{qp}$

$$\sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}x_p\bar{x}_q = \sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}\bar{x}_p x_q$$

Now the last sum is the conjugate of the first and since they are equal they are real. Let $x_p = t_p + \sqrt{-1}s_p$, then

$$\sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}x_p\bar{x}_q = \sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}t_p t_q + \sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}s_p s_q + \sqrt{-1}\sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}(t_q s_p - t_p s_q)$$

But the latter sum is zero. Consequently

$$\sum_{p=1}^{n}\sum_{q=1}^{n} a_{pq}x_p\bar{x}_q \leqq 0$$

Similarly

$$\sum_{p=1}^{n}\sum_{q=1}^{n} d_{pq}x_p\bar{x}_q > 0$$

Consequently $z \geqq 0$

THEOREM 2. *The roots of the characteristic equation are all real and positive.*

*Proof*: If we write the characteristics equation in the form (6) we have

$$\sum_{p=1}^{mn}\sum_{q=1}^{mn} d_{pq}x_p x_q = x_1^2 + x_2^2 + \cdots + x_{mn}^2$$

and

$$\sum_{p=1}^{mn}\sum_{q=1}^{mn} a_{pq}x_p x_q = -\sum_{p=1}^{mn}\sum_{q=1}^{mn} b(i,j)(g_{pq}x_p - s_{pq}x_q)^2 - \sum_{p=1}^{mn+1}\sum_{q=1}^{mn+1} k(i,j)(c_{pq}x_p - y_{pq}x_q)^2$$

where each of $g_{pq}$, $s_{pq}$, $c_{pq}$, $y_{pq}$ is 0 or 1. An inspection of the characteristic determinant shows these formulas. The non-negative character of the roots then follows as a result of the lemma just proved. It can be shown in a variety of ways that roots are actually positive. An easy way is to suppose that $v_1(i, j)$ is a characteristic function and that $\lambda = 0$. Since $v_1(i, j) \not\equiv 0$ it must have a maximum or minimum different from zero. Suppose a maximum. We mean by this that there must be at least one point $(i, j)$ such that $v_1(i, j) \geqq v_1(i - 1, j)$, $v_1(i + 1, j)$, $v_1(i, j - 1), v_1(i, j + 1)$ with the equality mark not holding in all instances. This is inconsistent with (5)

**3. The solution $V(i, j)$.** In order to continue our study we shall remove the requirement $v(m, n + 1) = 0$ but retain the other boundary conditions of (4). We thus require that $v(i, j)$ be zero at all points of the boundary of the fundamental rectangle except at the point $(m, n + 1)$. We solve the resulting linear equations. Write down the equations letting $(i, j)$ range over all interior points of the rectangle. We now have $mn$ homogeneous linear equations in $mn + 1$ unknowns. We shall consider the matrix of the coefficients. We arrange the equations and the unknowns in the order previously used except that $v(m, n + 1)$ will be put first in the arrangement of the unknowns. This matrix for the case $m = n = 3$ is to be viewed if one takes the matrix of the characteristic determinant of (5) and puts a first column consisting of $k(3, 3)$ followed by eight 0's. The values $v(i, j)$ sought will be proportional to the determinants formed from this matrix by striking successive columns of the matrix and alternating sign. For definiteness we shall let $w(i, j)$ be precisely these determinants with alternating signs prefixed. The sign prefixed to $w(m, n + 1)$ is to be plus. Now $w(i, j)$ is a solution of equation (3) and satisfies the boundary conditions (4), except possibly that $w(m, n + 1) \not\equiv 0$. We notice that the determinant obtained for $w(m, n + 1)$ is the characteristic determinant.

THEOREM 3. *The function $w(i, j)$ is a polynomial in $\lambda$ of the following form*

$$(-1)^{m+n-i-j+1}B\lambda^{mn-m-n+i+j-1} + A(\lambda)$$

*where $A(\lambda)$ is of lesser degree in $\lambda$ and $B > 0$.*

*Proof:* We shall develop the determinant for $w(1, 2)$ formed from the matric for $m = n = 3$ as above described(see chart I). This will serve to illustrate a general procedure.

We develop this determinant by minors of those columns which contain no $\lambda$ so long as this is possible. First we follow the principal diagonal and obtain $k(3, 3)b(2, 3)b(13)k(12)\lambda^5 +$ (terms of less degree in $\lambda$). Consider some other succession of minors. For example consider $k(3, 2)$ times its cofactor. We obtain this cofactor by striking the (3, 3) column and the (3, 2) row and attaching the proper sign. Now choose an element in the (3, 2) column. We choose $b(2, 2)$ and form the product $k(3, 3)k(3, 2)b(2, 2)$ multiplied by the cofactor of $b(2, 2)$ in the determinant just obtained. We now proceed to the (2, 2) column. We choose $k(2, 2)$ noticing that to obtain its cofactor we do not strike a row that contains $\lambda$. We have now $k(3, 3)k(3, 2)b(2, 2)k(2, 2)$ multiplied by a determinant of the fifth order containing five $\lambda$'s no one of which is in the same row or col-

| | (1, 4) | (3, 3) | (2, 3) | (1, 3) | (1, 1) | (3, 2) | (2, 2) | (3, 1) | (2, 1) |
|---|---|---|---|---|---|---|---|---|---|
| (3, 3) | $k(3, 3)$ | $\lambda + R(3, 3)$ | $b(2, 3)$ | 0 | 0 | $k(3, 2)$ | 0 | 0 | 0 |
| (2, 3) | 0 | $b(2, 3)$ | $\lambda + R(2, 3)$ | $b(1, 3)$ | 0 | 0 | $k(2, 2)$ | 0 | 0 |
| (1, 3) | 0 | 0 | $b(1, 3)$ | $\lambda + R(1, 3)$ | 0 | 0 | 0 | 0 | 0 |
| (1, 2) | 0 | 0 | 0 | $k(1, 2)$ | $k(1, 1)$ | 0 | $b(1, 2)$ | 0 | 0 |
| (1, 1) | 0 | 0 | 0 | 0 | $\lambda + R(1, 1)$ | 0 | 0 | 0 | $b(1, 1)$ |
| (3, 2) | 0 | $k(3, 2)$ | 0 | 0 | 0 | $\lambda + R(3, 2)$ | $b(2, 2)$ | $k(3, 1)$ | 0 |
| (2, 2) | 0 | 0 | $k(2, 2)$ | 0 | 0 | $b(2, 2)$ | $\lambda + R(2, 2)$ | 0 | $k(2, 1)$ |
| (3, 1) | 0 | 0 | 0 | 0 | 0 | $k(3, 1)$ | 0 | $\lambda + R(3, 1)$ | $b(2, 1)$ |
| (2, 1) | 0 | 0 | 0 | 0 | $b(1, 1)$ | 0 | $k(2, 1)$ | $b(2, 1)$ | $\lambda + R(2, 1)$ |

$$= k(3, 3)b(2, 3)b(1, 3)k(1, 2)$$

| | | | | |
|---|---|---|---|---|
| $\lambda + R(1, 1)$ | 0 | 0 | 0 | $b(1, 1)$ |
| 0 | $\lambda + R(3, 2)$ | $b(2, 2)$ | $k(3, 1)$ | 0 |
| 0 | $b(2, 2)$ | $\lambda + R(2, 2)$ | 0 | $k(2, 1)$ |
| 0 | $k(3, 1)$ | 0 | $\lambda + R(3, 1)$ | $b(2, 1)$ |
| $b(1, 1)$ | 0 | $k(2, 1)$ | $b(2, 1)$ | $\lambda + R(2, 1)$ |

$$+ k(3, 3)k(3, 2)b(2, 2)b(1, 2)$$

| | | | | |
|---|---|---|---|---|
| $\lambda + R(2, 3)$ | $b(1, 3)$ | 0 | 0 | 0 |
| $b(1, 3)$ | $\lambda + R(1, 3)$ | 0 | 0 | 0 |
| 0 | 0 | $\lambda + R(1, 1)$ | 0 | $b(1, 1)$ |
| 0 | 0 | 0 | $\lambda + R(3, 1)$ | $b(2, 1)$ |
| 0 | 0 | $b(1, 1)$ | $b(2, 1)$ | $\lambda + R(2, 1)$ |

Plus additional terms as to be explained.

CHART I

umn as any other. This succession of minors has consequently lead us to the product $\pm k(3, 3)k(3, 2)b(2, 2)k(2, 2)\lambda^5$. As a matter of fact the term is positive. This follows from the fact that $k(3, 3)k(3, 2)$, $b(2, 2)$, $k(2, 2)$, $\lambda, \lambda, \lambda, \lambda, \lambda$ which we have multiplied together can be put into the principal diagonal by interchanges of rows and of columns as now to be explained. The explanation is in general terms: We put the elements in question into the principal diagonal of the characteristic determinant by a rearrangement of the number pairs $(i, j)$. The determinant for $v(p, q)$ and its rearrangement are gotten by striking from the original characteristic determinant and its rearrangement the $(m, n)$ row and the $(p, q)$ column. Hence the determinant that we have is the minor of the element in the $(m, n)$ row and $(p, q)$ column. Now every element in the characteristic determinant is put into its new position by the interchange of rows with rows and a like interchange of columns. Such an operation on any determinant leaves it unchanged in value. If the determinants are expanded by minors of the same row in its old position and in its new position we have an identity. Hence, the minor of an element in one position must equal the minor of that element in its new position. Consequently the minors in question in our problem, namely from the original characteristic determinant and from its rearrangement are equal. Hence the sign of our product is positive.

Let us temporarily return to the $m = n = 3$ case. In the expansion of $w(1, 2)$ each term containing $\lambda^5$ corresponds in the rectangular figure to a path from (3, 4) to (3, 3) to (1, 2) by the smallest number of steps namely 4. This, (3, 4) to (3, 3) to (3, 2) to (2, 2) to (1, 2) corresponds to $k(3, 4)k(3, 3)b(2, 2)k(2, 2)\lambda^5$. The total coefficient of $\lambda^5$ is the sum of such products. A large number of steps will result in a product of less degree in $\lambda$. In general the coefficient of the highest power of $\lambda$ in the expansion of $w(p, q)$ is the sum of products obtained by following paths from $(m, n)$ to $(p, q)$ each step being either downward or to the left. (When a step is to the left a factor involving $b$ is obtained. Thus if the path leads from $(i, j)$ to $(i - 1, j)$ the factor $b(i, j)$ is introduced. If the path leads downward to $(i, j - 1)$ the factor $k(i, j)$ is introduced). The degree of $w(m, n)$ is by inspection $mn$. The degree of $w(i, j)$ is $mn$ minus the above minimum number of steps, namely, $mn - (m - i) - (n - j) - 1$.

We remove the highest common factor from $w(i, j)$, $i = 1, \cdots, m$; $j = 1, \cdots, n$. We now have a solution which we denote by $V(i, j)$ which is identically zero in $(i, j)$ for no value of $\lambda$ and which satisfies the conditions of being zero at all points of the boundary of our fundamental rectangle unless it is at $(m, n + 1)$. Let us suppose that the highest common factor which we have just removed is of degree $p$. Now the highest power of $\lambda$ in $w(i, j)$ is $mn - (m - i) - (n - j) - 1$. This is at least $mn - n - m + 1$. Hence, $p \leqq mn - n - m + 1$.

We now state the following theorem based on theorem II.

THEOREM 4. *The solution $V(i, j)$ is a polynomial in $\lambda$ of the form*

$$(-1)^{m+n-i-j+1}\bar{B}\lambda^{mn-m-n+i+j-p-1} + \bar{A}(\lambda)$$

*where $\bar{A}(\lambda)$ is of lesser degree in $\lambda$ and $p \leqq mn - m - n + 1$ and $\bar{B} > 0$, and where $p$ is independent of $i, j$.*

A second and immediate result is the following theorem

THEOREM 5. *If $\lambda > 0$ is sufficiently large $V(i, j)$ alternates in sign over the interior lattice arrangement of our fundamental rectangle. If $\lambda < 0$ is sufficiently large numerically $V(i, j)$ retains the same sign over the interior lattice arrangement of the fundamental rectangle.*

**4. Nodal Lines and the Zeros of $V(m, n + 1)$.** We make a plot of $z = V(i, j)$ by erecting at points $(i, j)$ of the fundamental rectangle $Z$-coordinates equal to $V(i, j)$. This constitutes a complete graph of the function $z$. However, we connect the points $[i, j, V(i, j)]$, $[I + 1, j, V(i + 1, j)]$ and $[i, j + 1, V(i, j + 1)]$ also $[i + 1, j, V(i + 1, j)][i, j + 1, V(i, j + 1)]$ and $[i + 1, j + 1, V(i + 1, j + 1)]$ by plane triangular regions. These triangles form a surface over our fundamental rectangle which we shall call the *roof*.

THEOREM 6. *The roof never meets the horizontal plane ($z = 0$) in an isolated point.*

This would be inconsistent with equation (5). Assume $V(i, j) = 0$. Then it is impossible for $V(i - 1, j)V(i + 1, j)$, $V(i, j - 1)$, $V(i, j + 1)$ all to have the same sign and have (5) satisfied.

*A point in which a straight line connecting $[i, j, V(i, j)]$ with $[i + 1, j, V(i + 1, j)]$* or with $[i, j + 1, V(i, j + 1)]$ *intersects the horizontal plane is called a node.*

*A line in which the roof cuts the horizontal plane is called a nodal line.*

THEOREM 7. *When $\lambda$ varies no nodal line on the fundamental rectangle ever lifts out in whole or in part nor does any nodal line enter by the reverse process.*

*Proof:* No nodal line ever lifts out in whole or in part because in order for this to happen it would be necessary to have points at which $V(i, j)$ is zero completely surrounded by points at which $V(i, j)$ is of fixed sign or zero but not all zero. This is inconsistent with equation (5). For the same reason no new nodal line can enter by the reverse process. It is well to remark again at this juncture that $V(i, j)$ is never identically zero in $(i, j)$. The highest common factor in $\lambda$ was removed from $w(i, j)$ to form $V(i, j)$. By the same means we can show the truth of the following two theorems:

THEOREM 8. *Every nodal line either crosses the fundamental rectangle from boundary to boundary or is a closed curve within the fundamental rectangle.*

THEOREM 9. *No nodal line ever disappears by complete fusion with another nodal line.*

As an immediate consequence of the previous theorems we have the following theorem:

THEOREM 10. *The only way that the number of nodal lines on the fundamental rectangle can decrease when $\lambda$ decreases is by exit of a nodal line at $(m, n + 1)$.*

We next prove the following theorem.

THEOREM 11. *When $\lambda$ decreases any node of $V(i, j)$ on the interval $n < y < n + 1$ moves continuously upward.*

*Proof:* Let $\bar{\lambda} = \lambda + \Delta\lambda$ where $\Delta\lambda > 0$ and denote the corresponding functions by a dash thus $\bar{V}(i, j)$, we then readily verify the following identity.

$$\Delta_i\{b(i-1,j)V(i-1,j)\Delta_i\bar{V}(i-1,j) - b(i-1,j)\bar{V}(i-1,j)\Delta_iV(i-1,j)\}$$

$$+ \Delta_j\{k(i,j-1)V(i,j-1)\Delta_j\bar{V}(i,j-1) \tag{8}$$

$$- k(i,j-1)\bar{V}(i,j-1)\Delta_jV(i,j-1)\} = -V(i,j)\bar{V}(i,j)\Delta\lambda$$

Apply the operator $\sum_{i=1}^{m}\sum_{j=1}^{n}$ or the equivalent $\sum_{j=1}^{n}\sum_{i=1}^{m}$ to both sides of (8). Let $|\Delta\lambda|$ be so small that $V(i,j)\ \bar{V}(i,j) > 0$ if $V(i,j) \neq 0$. We arrive at

$$V(m,n)\Delta_j\bar{V}(m,n) - \bar{V}(m,n)\Delta_jV(m,n) > 0 \tag{9}$$

We note that we have the strict inequality since $V(i,j) \not\equiv 0$ in $i, j$. We note from (9) that if $|\Delta\lambda|$ is sufficiently small $V(m,n)$ and $V(m,n+1)$ are not simultaneously zero.

The theorem is immediately drawn from (9) with the knowledge that all the functions involved are polynomials in $\lambda$.

We have noted that when $\lambda$ is very large the $V(i,j)$ alternate in sign and that when $\lambda$ is negative and numerically large they are all of the same sign. We note that when $\lambda$ is positive and large the number of nodal lines is $m+n-1$, and that when $\lambda$ is negative and numerically large the number of nodal lines is zero. We also recall that all characteristic values are positive. We now conclude theorem XI.

THEOREM 12. *All characteristic values are positive and these exist at least* $m+n-1$ *distinct characteristic value of* $\lambda$ *which we denote by* $\lambda_p$, $p = 0, 1, \cdots, m+n-2$, such that

$$0 < \lambda_0 < \cdots < \lambda_{m+n-3} < \lambda_{m+n-2}$$

*and such that when* $\lambda = \lambda_p$ *then* $V(i,j)$ *is a characteristic function and it has at least* $p$ *nodal lines on the fundamental rectangle. All other characteristic values are roots of the greatest common devisor of* $w(i,j)$, $i = 1, \cdots m$. $j = 1, \cdots, n$.

In the count of nodal lines if, when $\lambda$ varies, two nodal lines should meet at one or more points they are still counted as two nodal lines. It has been pointed out that two nodal lines never completely amalgamate.

**5. The Symmetrical Square.** Suppose $m = n$ and $b(i,j) = k(j,i)$.

THEOREM 13. *Under the above hypotheses* $w(i,j) = w(j,i)$.

*Proof:* We recall that $w(i,j)$ is given by a determinant formed from the characteristic determinant by striking out the $(m,n)$ row and the $(i,j)$ column multiplying by $k(m,n)$ and then prefixing the proper sign, positive if $(i,j)$ lies immediately to the right of $(m,n)$ and alternating thereafter. Rewrite the characteristic determinant with an interchange of $i$ and $j$ throughout. This amounts to an interchange of rows and a like interchange of columns. It does not alter the value of the characteristic determinant. Now in place of $b(i,j)$ write $k(j,i)$ and in place of $k(i,j)$ write $b(j,i)$. The characteristic determinant is now in its original form. Moreover, the position of $(j,i)$ relative to $(m,n)$ is the same as was the position of $(i,j)$ previous to the interchanges.

It is suggested that the reader carry out the above indicated steps on the characteristic determinant for $m = n = 3$ as exhibited in §2.

We now give several simple theorems.

THEOREM 14. *If $y(i, j)$ is a solution of (5) satisfying the conditions $y(0, i) = y(i, 0) = y(i, i) = y(i - 1, i) = y(i, i - 1) = 0$, $i = 1, \cdots, n$, then $y(i, j) \equiv 0$ throughout the fundamental square, $1 \leqq i \leqq n$, $1 \leqq j \leqq n$.*

*Proof:* The square can be filled in by use of equation (5) point by point beginning with the lower left hand corner. It is immediately seen that all values are zero.

THEOREM 15. *$V(i, i)$, $V(i - 1, i)V(i, i - 1)$ are not all zero, $1 \leqq i \leqq n$.*

*Proof:* The theorem follows from the fact that $V(i, j) \not\equiv 0$ over the fundamental square.

Let us now consider the diagonal of the fundamental square from lower left to upper right. On this diagonal are located the points (1, 1), (2, 2), $\cdots$. We consider the sequence of points on this line with coordinates $(\frac{1}{2}, \frac{1}{2})$, (1, 1), $(1 + \frac{1}{2}, 1 + \frac{1}{2})$, (2, 2), $(2 + \frac{1}{2}, 2 + \frac{1}{2})$, $\cdots$ $(n, n)$, $(n + \frac{1}{2}, n + \frac{1}{2})$. These constitute a set of $(2n + 1)$ equally spaced points. We denote them by $p = 0, 1, \cdots, 2n$. We define $w(i + \frac{1}{2}, i + \frac{1}{2})$ to equal $w(i + 1, i)$ which equals $w(i, i + 1)$. We now let

$$\tilde{w}(p) = w\left(\frac{p + 1}{2}, \quad \frac{p + 1}{2}\right)$$

A similar definition is given to $\tilde{V}(p)$ and to $\tilde{b}(p) = \tilde{k}(p)$.

It follows from (5) and the symmetry of $w(i, j)$ and of $V(i, j)$ that both $\tilde{w}(p)$ and $\tilde{V}(p)$ satisfy the equation,

$$\Delta\{\tilde{b}(p)\Delta y(p)\} + \tfrac{1}{2}\lambda y(p + 1) = 0 \tag{10}$$

This is a Sturm-Liouville normal equation in one variable. A fairly complete theory for it is known*.

THEOREM 16. $\tilde{V}(1) \not\equiv 0$

*Proof:* If $\tilde{V}(1) = 0$, since $\tilde{V}(0) = 0$, then $\tilde{V}(p) \equiv 0$, $p = 0, \cdots, 2n$. If $\tilde{V}(p) \equiv 0$ then $V(i, j) \equiv 0$ over the fundamental square. This is contrary to the fact.

For definiteness we assume $\tilde{V}(1) = V(1, 1) > 0$.

THEOREM 17. *The roots of $w(1, 1)$ are all positive.*

*Proof:* Expand the determinant for $w(1, 1)$ in terms of the minors of the (1, 1) row. We have $[k(1, 1) + b(1, 1)] \cdot D$ where $D$ is a determinant of the type treated in lemma 1.

THEOREM 18. *$V(1, 1)$ is a constant*

*Proof:* $V(1, 1) = \tilde{V}(1)$ is the result of removing a polynomial factor from $w(1, 1)$. It consequently is a real polynomial with no imaginary zeros that does not vanish. It consequently is a constant.

THEOREM 19. $\tilde{w}(j) = c\tilde{w}(1)V(j) = cw(1, 1)\tilde{V}(j)$

*Proof:* This theorem follows from the readily proved fact that two solutions

* Fort, T.: Finite Differences and Difference Equations in the Real Domain. Chapter X.

of (10) both zero when $p = 0$ are proportional. Here $c$ is independent of $\lambda$, but $w(1, 1)$ is in general dependent upon $\lambda$.

*Corollary:*

$$w(n + 1, n) = cw(1, 1)V(n + 1, n)$$

Now it is known that as $\lambda$ decreases the nodes of $V(p)$ which may lie in the interval $1 < p < 2n$ move continuously from lower left to upper right. In other words nodal lines of $V(i, j)$ move continuously outward along this diagonal line. We are now in a position to state the following theorem.

THEOREM 20. *There are precisely $2n - 1$ positive characteristic values $\lambda_q$ such that when $\lambda = \lambda_q$ then $V(n + 1, n) = 0$ with precisely $q$ nodal lines crossing the diagonal running from the lower left to upper right corner of the fundamental square. Moreover $\lambda_{q+1} > \lambda_q > 0$*

We note that we have not proved that there are not other characteristic values for which $w(i, j) \equiv 0$. For such a characteristic value, if there are such, there must be some other solution not identically zero; also such characteristic values must be zeros of $w(1, 1)$.

**6. The Factorable Case.** If $k(i, j) = M(i)N(j)$ and $b(i, j) = B(i)D(j)$ then equation (3) can be written

$$D(j)\Delta_i\{B(i - 1)\Delta_i y(i - 1, j)\} + M(i)\{\Delta_j N(j - 1)\Delta_j y(i, j - 1)\} + \lambda y(i, j) = 0. \tag{11}$$

If now we assume $y(ij) = s(i)r(j)$ and divide through by $y(i, j)D(j)M(i)$ we get

$$\frac{\Delta_i\{B(i - 1)\Delta_i s(i - 1)\}}{s(i)} + \frac{\Delta_j\{N(j - 1)\Delta_j r(j - 1)\}}{r(j)} + \lambda = 0$$

Here we have the sum of a function of $i$ plus a function of $j$ equal to a constant. Each of the functions is consequently a constant. We consequently write

$$\Delta_i\{B(i - 1)\Delta_i s(i - 1)\} + Vs(i) = 0 \tag{12}$$

and

$$\Delta_j\{N(j - 1)\Delta_j r(j - 1)\} + \rho r(j) = 0 \tag{13}$$

where $v + \rho = \lambda$

If now $s(i)$ is a solution of the one dimensional Sturm-Liouville system consisting of (12) and the boundary conditions $s(0) = s(n + 1) = 0$; and if $r(j)$ is a solution of the one dimensional Sturm-Liouville system consisting of (13) and the boundary conditions $r(0) = r(n + 1) = 0$; then $r(i)s(i)$ is a solution of the two dimensional Sturm-Liouville system consisting of (11) together with the boundary conditions $y(i, 0) = y(i, n + 1) = y(0, j) = y(n + 1, j) = 0$ $i = 1, 2, \cdots, n$, $= 1, 2, \cdots, n$. The characteristic value is the sum of the respective characteristic values.

UNIVERSITY OF GEORGIA
OFFICE OF ORDNANCE RESEARCH OF THE UNITED STATES ARMY

(Received June 8, 1953)

# THE CONVERGENCE OF THE SOLUTIONS OF A CLASS OF ITERATIVE DIFFERENCE EQUATIONS

By J. Samoloff

**Introduction.** This note stems from a study of Iterative processes involving Mean Value functions. There is an extensive literature on General Mean functions [1], and on Iterations [2], [3] (G. Jessen, K. Knopp, H. Geppert, G. Aumann, etc.). The condition of positive Separation, as introduced in the following paragraphs, is equivalent to one* of G. Auman's [4] conditions, although the present author arrived at the idea independently.

As used in this article, a (weak) Mean is a real valued function of real variables, whose functional values all lie between the greatest and the least of the independent variables. In symbols

$$\max (x_j) \geqq f(x_1 , \cdots x_p) \geqq \min (x_j) \tag{1}$$

Note that this condition immediately implies that

$$\lim_{\text{all } x_j \to x} f(x_1 , \cdots x_p) \equiv x \tag{2}$$

Henceforth we shall assume (or define)

$$f(x, x, \cdots x) \equiv x \tag{3}$$

if $f$ is a mean.

**The Problem.** We shall introduce and apply a general result concerning the convergence of the special iteration process afforded by the $p^{\text{th}}$ order difference equation

$$a_{n+p} = f(a_n , a_{n+1} , \cdots a_{n+p-1}), \qquad n = 0, 1, \cdots \tag{4}$$

with prescribed initial values $a_0 , a_1 , \cdots a_{p-1}$. As a simple example, the sequence $\{a_n\}$ defined by $a_{n+2} = \frac{1}{2}(a_n + a_{n+1})$ converges for all choices of $a_0 , a_1$, to the function $\frac{1}{3}(a_0 + 2a_1)$. This is a special case of example 1, discussed later.

We notice that if $f$ is a mean satisfying (1) and the $p$-tuple of initial values is selected from a bounded region $C$ defined by the inequalities

$$A \geqq \text{all } (a_j) \geqq B \tag{5}$$

then all the members of the solution sequence $\{a_n\}$ will likewise stay in the region $C$. This means that the sequence must either converge or oscillate finitely. The introduction of the Separation conditions (12) on $f$ is one way of ensuring convergence, as we shall see.

* $\max (x_j) - f \geqq \rho(\max (x_j) - \min (x_j))$
$f - \min (x_j) \geqq \rho(\max (x_j) - \min (x_j))$ for some fixed $\rho > 0$.

**Separations of Means.** Let $f(P) = f(x_1, \cdots x_p)$ be a mean function of the real $p$-tuple $P = (x_1, \cdots x_p)$ defined on the region $C$ (5)*. Let

$$\delta = \max(x_j) - \min(x_j) \tag{6}$$

Then the Separation numbers $S$, $s$, (called the Upper Separation and Lower Separation, respectively, of $f$ on the region $C$), are defined by

$$\begin{aligned} S &= \operatorname*{Inf}_{\substack{\text{all } P \in C \\ \delta \neq 0}} \dagger \left\{ \frac{\max(x_j) - f}{\delta} \right\} \\ s &= \operatorname*{Inf}_{\substack{\text{all } P \in C \\ \delta \neq 0}} \left\{ \frac{f - \min(x_j)}{\delta} \right\} \end{aligned} \tag{7}$$

It is clear that for a mean, the numbers $S$ and $s$ are finite; in fact we immediately deduce

$$0 \leqq S \leqq 1, \qquad 0 \leqq s \leqq 1 \tag{8}$$

and

$$S + s \leqq 1 \tag{9}$$

Setting

$$S' = 1 - S, \qquad s' = 1 - s, \tag{10}$$

the following inequalities may be readily shown to hold:

$$S' \max(x_j) + S \min(x_j) \geqq f \geqq s \max(x_j) + s' \min(x_j) \tag{11}$$

For e.g.,

$$\begin{aligned} (1 - S) \max(x_j) + S \min(x_j) - f(P) &= \max(x_j) - f(P) - S\delta \\ &= \delta \left[ \left\{ \frac{\max(x_j) - f(P)}{\delta} \right\} - S \right] \geqq 0 \end{aligned}$$

by definition of $S$.

So far, our function need only satisfy (1). We now define a Strong Mean as a mean for which

$$S > 0, \qquad s > 0. \tag{12}$$

Then by virtue of (9)

$$0 < S < 1, \qquad 0 < s < 1. \tag{13}$$

Intuitively, this conveys the idea that the values of the function $f$ are to lie strictly between the greatest and least of its variables, being separated from them by at least a certain positive fraction of $\delta = \max(x_j) - \min(x_j)$, regardless of where in $C$ the interval $\delta$ may be placed.

* Thinking of $(x_1, \cdots x_p)$ as the coordinates of a point in Euclidean $p$-space, $C$ is a square in two dimensions, a cube in three dimensions, with the "straight line" $x_1 = x_2 = \cdots = x_p$ as a main diagonal.

† or Greatest Lower Bound (g.l.b.).

Examples of strong means are the classical functions

$$\mathfrak{M}_\phi = \phi^{-1}\{\textstyle\sum_{j=1}^{p} w_j \phi(x_j)\} \tag{14}$$

where $w_j > 0$, $\sum_1^p w_j = 1$, and $\phi$ is any (strictly) monotonic function, $\phi^{-1}$ being its inverse.

Many functions not of this form, but satisfying (2), are strong means if the region $C$ is suitably chosen: e.g.

$$f(x_1, x_2) = x_2^2 + x_2 - x_1^2 \tag{15}$$

is a strong mean for

$$0 > A \geqq \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geqq B > -\tfrac{1}{2}. \tag{16}$$

For if $x_1 > x_2$,

$$\frac{x_1 - f}{x_1 - x_2} = \frac{x_1 - x_2 + x_1^2 - x_2^2}{x_1 - x_2} = 1 + x_1 + x_2$$

and

$$\frac{f - x_2}{x_1 - x_2} = \frac{x_2^2 - x_1^2}{x_1 - x_2} = -(x_1 + x_2)$$

Hence for positive separations the points $(x_1, x_2)$ must lie between the lines $x_1 + x_2 = 0$ and $x_1 + x_2 = -1$, and since the region $C$ in two dimensions is a square with the line $x_1 = x_2$ as a diagonal, it must be the region specified by the inequalities (16).

It should be noticed that the distinction between weak and strong means, although essential to our proofs to follow, is really a matter of restricting the region $C$. For instance, the last example (15) becomes a weak mean if we allow $A = 0$, $B = -\frac{1}{2}$.

We shall now state our theorem, give a few applications and conclude with the proof. The proof, incidentally, employs only the most elementary ideas of analysis, but perhaps in a somewhat detailed manner.

**Theorem.** Let $f(x_1, \cdots x_p)$ be a strong mean on $C$: $A \geqq x_j \geqq B$; then the solution sequence $\{a_n\}$ of the difference equation

$$a_{n+p} = f(a_n, a_{n+1}, \cdots a_{n+p-1})$$

converges uniformly for all sets of initial values

$$(a_0, a_1, \cdots, a_{p-1}) \varepsilon C$$

to a function $F(a_0, \cdots a_{p-1})$ which is uniquely defined by the properties

(a) $\lim_{\text{all } a_j \to x} F(a_0, \cdots a_{p-1}) \equiv x$

(b) $F(a_0, a_1, \cdots a_{p-1}) \equiv F[a_1, a_2, \cdots a_{p-1}, f(a_0, \cdots a_{p-1})]$

COROLLARIES:

(i) $F(a_0, \cdots a_{p-1})$ is continuous on $C$ if $f(a_0, \cdots a_{p-1})$ is.

(ii) If $f$ is homogeneous (of degree one), so is $F$.

(iii) $F(a_0, \cdots a_{p-1})$ is a mean on $C$.

**Examples.** In the following examples the use of the theorem enables us to conclude that the solution sequence $\{a_n\}$ converges uniformly to a continuous function $F(a_0, \cdots a_{p-1})$, and to establish the formula for $F$ by using the uniqueness conditions (a) and (b).

1. If $a_{n+p} = \{\sum_{j=0}^{p-1} w_j a_{j+n}^r\}^{1/r}$, $n = 0, 1, \cdots$ where $w_j > 0$, $\sum_1^{p-1} w_j = 1$, $a_j > 0, j = 0, 1, \cdots p-1$.

Then $\lim_{n\to\infty} \{a_n\}$ exists and is equal to

$$\{1/v \sum_{j=0}^{p-1} (\sum_{i=0}^{j} w_i)\, a_j^r\}^{1/r}$$

where $v = \sum_0^{p-1} (p-j)\, w_j$. For as is easily verified, the proposed expression satisfies the uniqueness conditions (a) and (b) of the Theorem. Explicit expressions for $F(a_0 \cdots a_{p-1})$ can, however, only be obtained in special cases.

To show how to proceed in attempting to apply the Theorem to a given equation of the form (4), where $f$ is not a Mean, we adduce example

2. $a_{n+1} = a_n^2 - k$.

It will turn out that in order to apply the theorem, we shall have to restrict $k$ as well as the initial value $a_0$. We write

$$\begin{aligned} a_{n+2} &= a_{n+1}^2 - k = a_{n+1}^2 + a_{n+1} - a_n^2 \\ &= f(a_n, a_{n+1}) \end{aligned}$$

and $f(x_1 x_2) = -x_1^2 + x_2 + x_2^2$ is, as we have seen, a Strong Mean for $0 > \dfrac{x_1}{x_2} > -\frac{1}{2}$.

Hence the solution sequence converges if we choose $-\frac{1}{2} < \dfrac{a_0}{a_1} < 0$. But $k = -a_1 + a_0^2$. Hence we must choose $-\frac{1}{2} < a_0 < 0$, $0 < k < \frac{3}{4}$.

It can be verified by other methods that the example converges for the stated values of $k$ and $a_0$. In this case the limit is

$$\tfrac{1}{2}\{1 - \sqrt{1 + 4k}\} = \tfrac{1}{2}\{1 - \sqrt{1 + 4(a_0^2 - a_1)}\} = F(a_0, a_1)$$

and it is easily verified that this satisfies the uniqueness conditions.

In general, to apply the theorem to

$$a_{n+p} = f(a_n, \cdots a_{n+p-1}) - k$$

where the R.H.S. does not satisfy (2), if $f$ is a continuous function, we write

$$\begin{aligned} a_{n+p+1} &= f(a_{n+1} \cdots a_{n+p}) + a_{n+p} - f(a_n \cdots a_{n+p-1}) \\ &= \phi(a_{n+1} \cdots a_{n+p}) \end{aligned}$$

and $\phi$ satisfies 2. Next we investigate whether $C$ exists such that $\phi$ is a strong Mean on $C$, and choose $a_0, a_1 \cdots a_{p-1}$ and $k$ accordingly.

**Proofs.** We shall first prove the convergence of $\{a_n\}$.

Let $S, s$ be the separations of $f$ on $C$.

From the inequality (1) we conclude that

$$\max(a_m, \cdots a_{m+p-1}) - \min(a_m, \cdots a_{m+p-1}) \leqq \max(a_n, \cdots a_{n+p-1}) - \min(a_n, \cdots a_{n+p-1}) \tag{A}$$

if $m \geqq n \geqq 0$.

Let $a_{n+g} = \max(a_i)$, $a_{n+l} = \min(a_i)$, for $n \leqq i \leqq n+p-1$.

Then $0 \leqq g \leqq p-1$, $0 \leqq l \leqq p-1$. If $g = l$, the proof is trivial, by the remark (A). Suppose $g > l$. Then using the inequalities (11)

$$\begin{aligned} a_{n+g} - a_{n+l} &= f(a_{n+g-p} \cdots a_{n+g-1}) - a_{n+l} \\ &\leqq S' \max(a_j) + S \min(a_j) - \min(a_j) \\ &= S'[\max(a_j) - \min(a_j)], \qquad n+g-p \leqq j \leqq n+g-1 \\ &\leqq S'[\max(a_{j'}) - \min(a_{j'})], \qquad n-p \leqq j' \leqq n-1 \end{aligned}$$

by (A). Similarly, if $g < l$,

$$\begin{aligned} a_{n+g} - a_{n+l} &= a_{n+g} - f(a_{n+l-p}, \cdots a_{n+l-1}) \\ &\leqq s'[\max(a_{j'}) - \min(a_{j'})], \qquad n-p \leqq j' \leqq n-1. \end{aligned}$$

In either case, letting $\rho = \max(S', s')$,

$$\begin{aligned} 0 &\leqq \max(a_i) - \min(a_i), && n \leqq i \leqq n+p-1 \\ &\leqq \rho\ [\max(a_{j'}) - \min(a_{j'})], && n-p \leqq j' \leqq n-1 \\ &< \rho^m[\max(a_k) - \min(a_k)], && 0 \leqq n-mp \leqq k \leqq n-(m-1)p-1 \\ &< \rho^m[\max(a_{k'}) - \min(a_{k'})], && 0 \leqq k' \leqq p-1 \end{aligned}$$

where $0 < \rho < 1$, and

$$m = [n/p] = \text{greatest integer} \leqq n/p.$$

Finally, $0 \leqq \max(a_i) - \min(a_i) \leqq \rho^m \mid A - B \mid$ for all $i$ such that $n \leqq i \leqq n+t$, and arbitrary $t > 0$, by (A).

This proves the uniform convergence of $\{a_n\}$ on $C$, since $\rho$ is independent of $a_0, \cdots a_{p-1}$.

The necessity of the conditions (a) and (b) on $F(a_0, \cdots a_{p-1})$ now follows as a corollary.

To prove the sufficiency, let $\phi(a_0, \cdots a_{p-1})$ be any function satisfying (a) and (b). By property (b) we notice that

$$\begin{aligned} \phi(a_n, \cdots a_{n+p-1}) &\equiv \phi\{a_{n+1}, \cdots a_{n+p-1}, f(a_n, \cdots a_{n+p-1})\} \\ &\equiv \phi\{a_{n+1}, \cdots a_{n+p}\} \end{aligned}$$

where $\{a_n\}$ satisfies (4).

Since this is true for all $n \geqq 0$,

$$\phi(a_0 \cdots a_{p-1}) = \phi(a_1, \cdots a_p) = \cdots = \phi(a_n, \cdots a_{n+p-1})$$

and since $\lim_{n\to\infty} a_n = F(a_0 \cdots a_{p-1})$, it follows from condition (a) that

$$\phi(a_0 \cdots a_{p-1}) = F(a_0 \cdots a_{p-1})$$

This completes the proof.

## REFERENCES

[1] HARDY, LITTLEWOOD AND POLYA, "Inequalities."
[2] H. GEPPERT, "Iterative Algorithmen, I, II" Math. Annalen 107, 387-399, 108, 197-207 (1932-33).
[3] G. Aumann, "Über den Mischalgorithmus bei Analytischen Mittelwerten," Math. Zeit. 39, 625-629 (1935).
[4] G. AUMANN, "Aufbau von Mittelwerten Mehrerer Argumente, I, II", Math. Annalen 109, 235-253 (1933), 111, 713-730 (1935).

LEHIGH UNIVERSITY

(Received February 6, 1953)

# REGULAR AND IRREGULAR COULOMB WAVE FUNCTIONS EXPRESSED IN TERMS OF BESSEL-CLIFFORD FUNCTIONS*

By Milton Abramowitz

**1. Statement of Problem.** Coulomb wave functions are defined as the solutions of the differential equation

$$\frac{d^2y}{d\rho^2} + \left[1 - \frac{2\eta}{\rho} - \frac{L(L+1)}{\rho^2}\right] y = 0 \tag{1.1}$$

where $L$ is a positive integer and $\eta$ is a continuous parameter defining the energy values. In the quantum mechanical applications the regular and irregular solutions of (1.1) identified as $F_L(\eta, \rho)$ and $G_L(\eta, \rho)$ are normalized so that as $\rho \to \infty$

$$G_L(\eta, \rho) + iF_L(\eta, \rho) \sim \exp\,[i(\rho - \eta \ln 2\rho - L\pi/2 + \sigma_L)] \tag{1.2}$$

where $\sigma_L = \arg \Gamma(L + 1 + i\eta)$.

For the past several years the Computation Laboratory of the National Applied Mathematics Laboratories of the National Bureau of Standards has been actively engaged in tabulating the solutions of (1.1). Thus far, there has appeared the first volume [1] and the second is now nearing publication. In addition there has been lately developed a comprehensive program for carrying out the computations on the National Bureau of Standards Eastern Automatic Computer (SEAC). According to present plan it will be possible to tabulate the functions $F_L$ and $G_L$ for $L = 0(1)25$ and $\eta$ and $\rho$ in the range from 0 to 100.

Although the results obtained here will be generalized for any $L$, for the sake of simplicity we shall confine the discussion to the case $L = 0$. The explicit expression for $F_0$ is

$$F_0(\eta, \rho) = C_0(\eta)\rho\Phi_0(\eta, \rho) \tag{1.3}$$

where

$$C_0^2(\eta) = 2\pi\eta(e^{2\pi\eta} - 1)^{-1} \tag{1.4}$$

and

$$\Phi_0(\eta, \rho) = \sum_{n=1}^{\infty} A_n\rho^{n-1} \tag{1.5}$$

and where the coefficients $A_n$ are given by

$$A_1 = 1, \quad A_2 = \eta, \quad n(n-1)A_n = 2\eta A_{n-1} - A_{n-2} \quad \text{for } n > 2. \tag{1.6}$$

The explicit expression for $G_0(\eta, \rho)$ is

$$G_0(\eta, \rho) = D_0(\eta)\Theta_0(\eta, \rho); \quad D_0(\eta)C_0(\eta) = 1 \tag{1.7}$$

where

* Preparation of this paper was sponsored (in part) by the Flight Research Laboratory, Wright Air Development Center, USAF.

$$\Theta_0(\eta, \rho) = 2\eta\rho\Phi_0(\eta, \rho)[\ln 2\rho + \tfrac{1}{2}\eta^{-1}q_0(\eta)] + \psi_0(\eta, \rho). \tag{1.8}$$

In (1.8) the expressions for $q_0(\eta)/2\eta$ and $\psi_0(\eta, \rho)$ are

$$q_0(\eta)/2\eta = -1 + 2\gamma + \text{Rl.Pt}\,\{\Gamma'(1 + i\eta)/\Gamma(1 + i\eta)\} \tag{1.9}$$

where $\gamma = .5772156649\ldots$ is Euler's constant, and

$$\psi_0(\eta, \rho) = \sum_{n=0}^{\infty} a_n\rho^n, \tag{1.10}$$

the coefficients $a_n$ being defined as

$$\begin{aligned} &a_0 = 1, \qquad a_1 = 0, \qquad n(n-1)a_n \\ &\qquad = 2\eta a_{n-1} - a_{n-2} - (2n-1)2\eta A_n \qquad \text{for } n \geqslant 2 \end{aligned} \tag{1.11}$$

where the coefficients $A_n$ are given by (1.6). The general expressions for $F_L(\eta, \rho)$, and $G_L(\eta, \rho)$ may be found in [1].

While the expression for $F_0(\eta, \rho)$ is relatively simple and lends itself to calculation, the formula for $G_0(\eta, \rho)$ is unwieldy and provides numerous pitfalls to the computer. In the cases where $L \neq 0$ the expressions are so involved that the direct evaluation of $G_L(\eta, \rho)$ from its series representation becomes a formidable task. In particular we note that when $\eta$ is large with respect to $\rho$ the expression for $\Theta_0(\eta, \rho)$ usually results in the difference between two nearly equal terms and there is a considerable loss in significant figures. For example when $\eta = 10$, $\rho = 1$ we find

$$\psi_0(\eta, \rho) = -12940.54119, \qquad \rho\Phi_0(\eta, \rho) = 205.34038,$$
$$2\eta[\ln 2\rho + q_0(\eta)/2\eta] = 63.019955$$

and therefore $\Theta_0(\eta, \rho) = .0005$. In this case there has been a loss of eight significant figures. In the present article we shall develop an expression for $\Theta_0$ which overcomes this difficulty. We shall give a series expansion in terms of the functions $z^{n/2}K_n(2\sqrt{z})$ suitable for large $\eta$ and small $\rho$ in which it will be possible to tell immediately the proper number of figures to be carried in the calculation. Furthermore, the coefficients in this series have the property that they are the same ones required in the determination of $\rho\Phi_0(\eta, \rho)$. The results as given are not essentially new, having been discussed previously by Breit and others [2 to 6].

**2. Derivation of Series Expansions for $\rho\Phi_0(\eta, \rho)$ and $\Theta_0(\eta, \rho)$.**

If we make the substitution $2\eta\rho = t$ in (1.1) for the case $L = 0$ we get

$$ty'' + [(t/4\eta^2) - 1]y = 0. \tag{2.1}$$

When $\eta \to \infty$ this equation approaches $ty'' - y = 0$ for which the general solution is $y = \alpha t^{\frac{1}{2}}I_1(2\sqrt{t}) + \beta t^{\frac{1}{2}}K_1(2\sqrt{t})$. It is therefore plausible to express the solutions of (2.1) in terms of the functions $t^{n/2}I_n(2\sqrt{t})$ and $t^{n/2}K_n(2\sqrt{t})$. However, in order that we shall be concerned with only one set of recurrence relations and the final result can be expressed by a single formula let $u_n$ stand for either of the functions $t^{n/2}I_n(2\sqrt{t})$ or $\cos n\pi \cdot t^{n/2}K_n(2\sqrt{t})$. We then have

$$tu_n'' + (1 - n)u_n' - u_n = 0, \tag{2.2}$$

$$u_n' = u_{n-1}, \tag{2.3}$$

$$tu_n = (n + 1)u_{n+1} + u_{n+2}. \tag{2.4}$$

Now let $H_0$ stand for either $\rho\Phi_0(\eta, \rho)$ or $\Theta_0(\eta, \rho)$ and we write formally

$$H_0(\eta, \rho) = \lambda \sum_{n=1}^{\infty} b_n u_n(t); \qquad t = 2\eta\rho \tag{2.5}$$

with the understanding that $u_n \equiv t^{n/2}I_n(2\sqrt{t})$ when $H_0 = \rho\Phi_0(\eta, \rho)$ and $u_n \equiv \cos n\pi \cdot t^{n/2}K_n(2\sqrt{t})$ when $H_0(\eta, \rho) = \Theta_0(\eta, \rho)$. Here $\lambda$ is a constant to be chosen from the conditions at $\rho = 0$. From (1.5) above we note that when $\rho = 0$, $\rho\Phi_0 = 0$ and $(d/d\rho)(\rho\Phi_0) = 1$. Also from (1.8) and (1.10) when $\rho = 0$ we have $\Theta_0 = 1$. Furthermore we have for $t = 0$, $t^{n/2}I_n(2\sqrt{t}) = 0$, $(d/d\rho)(t^{\frac{1}{2}}I_1(t)) = 2\eta$, $\cos n\pi \cdot t^{n/2}K_n(t) = (-1)^n(n - 1)!/2$.

Now, substituting (2.5) in (2.1) we get

$$\sum_{n=1}^{\infty} b_n\{tu_n'' + (t/4\eta^2)u_n - u_n\} = 0 \tag{2.6}$$

Making use of the relations (2.2) to (2.4) we then get

$$\sum_{n=1}^{\infty} b_n\{(n - 1)u_{n-1} + u_n + [(n + 1)/4\eta^2]u_{n+1} + (1/4\eta^2)u_{n+2} - u_n\} = 0 \tag{2.7}$$

or

$$\sum_{n=1}^{\infty} b_n\{4\eta^2(n - 1)u_{n-1} + (n + 1)u_{n+1} + u_{n+2}\} = 0. \tag{2.8}$$

Setting the coefficients of $u_n$ equal to zero we get

$$4\eta^2 n b_{n+1} + n b_{n-1} + b_{n-2} = 0 \tag{2.9}$$

For $n = 1$ we get $b_2 = 0$ and for $n = 2$ we get $4\eta^2 b_3 + b_1 = 0$ so that all coefficients $b_n$ for $n > 2$ can be generated in terms of $b_1$ from (2.9). Since (2.1) is a linear differential equation we may choose $b_1 = 1$ and it remains to identify $\lambda$. For the regular solution since $(\rho\Phi_0)' = 1$ for $\rho = 0$ we find $\lambda = 1/(2\eta)$ and for the irregular solution since $\Theta_0 = 1$ for $\rho = 0$ we must have

$$\lambda \sum_{n=1}^{\infty} (-1)^n b_n \tfrac{1}{2}(n - 1)! = 1$$

Summarizing our results we then have

$$\rho\Phi_0(\eta, \rho) = (1/2\eta) \sum_{n=1}^{\infty} b_n t^{n/2} I_n(2\sqrt{t}), \qquad t = 2\eta\rho \tag{2.10}$$

$$\Theta_0(\eta, \rho) = \lambda \sum_{n=1}^{\infty} (-1)^n b_n t^{n/2} K_n(2\sqrt{t}), \qquad t = 2\eta\rho \tag{2.11}$$

where

$$b_1 = 1, \quad b_2 = 0, \quad 4\eta^2 n b_{n+1} + n b_{n-1} + b_{n-2} = 0, \quad (n > 2) \tag{2.12}$$

$$2\lambda^{-1} = \sum_{n=1}^{\infty} (-1)^n b_n (n - 1)!$$

We have further from (2.3)

$$(d/d\rho)(\rho\Phi_0) = \sum_{n=0}^{\infty} b_{n+1} t^{n/2} I_n(2\sqrt{t}) \tag{2.13}$$

$$(d/d\rho)(\Theta_0) = 2\eta\lambda \sum_{n=0}^{\infty} (-1)^n b_{n+1} t^{n/2} K_n(2\sqrt{t}) \tag{2.14}$$

The advantage of these expressions is obvious. We need only compute one set of coefficients for both the regular and irregular solutions. From the known behavior of the functions $I_n(z)$ and $K_n(z)$ we know how many figures to carry in the computation. For some purposes it will be sufficient to use only the first term in these expressions. Finally, it should be mentioned that should one need values of $F_L$ and $G_L$, for values other than $L = 0$, it is probably easier to use (2.10)–(2.14) (when $\eta \gg \rho$) to determine $F_0$ and $G_0$ and their derivatives and then generate the values for $L > 0$ by the recurrence relations [1]. Although in general there will be a loss in accuracy due to successive application of the recurrence formula, by starting with enough figures one may end up with sufficient accuracy in the values desired.

From the asymptotic expansion for $I_n(z)$ and $K_n(z)$ we have $I_n(z) \sim e^z/\sqrt{2\pi z}$ and $K_n(z) \sim \sqrt{\pi/2z}\, e^{-z}$ for $z$ sufficiently large. Replacing the Bessel functions in (2.10) and (2.11) by these expressions and keeping only the first terms in the series, we obtain since $b_1 = 1$,

$$\rho F_0(\eta, \rho) \sim \sqrt{1/8\pi\eta}(\rho/2\eta)^{\frac{1}{4}} e^{2\sqrt{2\eta\rho}} \tag{2.15}$$

$$\Theta_0(\eta, \rho) \sim \sqrt{\pi}(2\eta\rho)^{\frac{1}{4}} e^{-2\sqrt{2\eta\rho}} \tag{2.16}$$

**3. Convergence of (2.10) and (2.11).** In order to study the convergence of the series representation obtained, we must first obtain an estimate for the coefficients $b_n$ in (2.12).

Let us put $\tau = 4\eta^2$ and $b_n = \exp[\phi(n)]$. If we further write $\phi(n + a) = \phi(n) + a\phi'(n)$ neglecting the derivatives of higher order than the first we get

$$\tau n\alpha^3 + \alpha n + 1 = 0; \qquad \alpha = \exp[\phi'(n)], \tag{3.1}$$

and therefore

$$\alpha^2 = -(1/\tau) - (1/\tau\alpha n). \tag{3.2}$$

Thus, if $n$ is large the solution of this equation is approximately

$$\alpha \cong (-\tau)^{-\frac{1}{2}} \tag{3.3}$$

and we get ultimately that

$$|\, b_n \,| \sim (2\eta)^{-n} \tag{3.4}$$

From the series expansion for $I_n(t)$ we have

$$t^{n/2} I_n(2\sqrt{t}) = \sum_{s=0}^{\infty} \frac{t^{s+n}}{\Gamma(s+1)\Gamma(n+s+1)} < \frac{t^n}{n!} e^t \tag{3.5}$$

and therefore

$$|\, b_n t^{n/2} I_n(2\sqrt{t}) \,| < \beta_1 t^n e^t / n!(2\eta)^n \tag{3.6}$$

where $\beta_1$ is a constant independent of $n$. Thus (2.10) converges uniformly and absolutely for all finite values of $t = 2\eta\rho$.

In order to discuss the convergence of (2.11) we note first [7] that

$$K_n(2\sqrt{t}) = \tfrac{1}{2}\int_0^\infty e^{-\sqrt{t}(\xi+\xi^{-1})}\xi^{-n-1}\,d\xi \tag{3.7}$$

$$\lim_{n\to\infty}(-1)^n t^{n/2}K_n(2\sqrt{t}) = \tfrac{1}{2}(-1)^n(n-1)! \tag{3.8}$$

Now if $b_n t^{n/2}K_n(2\sqrt{t}) = v_n$ applying the ratio test for convergence we obtain with the aid of (3.4) and (3.8)

$$|\, v_{n+1}/v_n \,| \to \beta_2 n/2\eta \tag{3.9}$$

where $\beta_2$ is a constant independent of $n$. Since this quantity approaches infinity we conclude that (2.11) diverges. However we may show that the representation (2.11) is an asymptotic expansion. Thus, if $S_{n-1}$ represents the sum of the first $(n-1)$ terms, consider

$$\begin{aligned} |\,\Theta_0 - S_{n-1}\,| &= \left|\lambda \sum_{m=n}^{\infty}(-1)^m b_m t^{m/2}K_m(2\sqrt{t})\right| \\ &\leqq |\,\lambda\,| \sum_{m=n}^{\infty} |\, b_m\,| t^{m/2}K_m(2\sqrt{t}) \qquad (3.10) \\ &\leqq \frac{|\,\lambda\,|}{2}\sum_{m=n}^{\infty}|\, b_m\,| t^{m/2}\int_0^\infty e^{-\sqrt{t}(\xi+\xi^{-1})}\xi^{-m-1}\,d\xi \end{aligned}$$

Inverting the order of integration and summation we get with the help of (3.4)

$$|\,\Theta_0 - S_{n-1}\,| \leqq \frac{|\,\lambda\,|}{2}\int_0^\infty e^{-\sqrt{t}(\xi+\xi^{-1})}\sum_{m=n}^{\infty}\frac{t^{m/2}}{(2\eta)^m\xi^{m+1}}\,d\xi \tag{3.11}$$

The dominant contribution to the integral comes in the neighborhood of $\xi = 1$. With this assumption we can take $2\eta$ so large that $(t^{\frac{1}{2}}/2\eta\xi) < 1$ and we have a geometric progression which can be summed. Thus

$$|\,\Theta_0 - S_{n-1}\,| \leqq \frac{|\,\lambda\,|}{2}\int_0^\infty e^{-\sqrt{t}(\xi+\xi^{-1})}\frac{t^{n/2}}{(2\eta)^n\xi^n}\frac{d\xi}{\xi - (t^{\frac{1}{2}}/2\eta)} \tag{3.12}$$

In the limit as $(2\eta) \to \infty$ we therefore have

$$|\,\Theta_0 - S_{n-1}\,| \leqq |\,\lambda\,|\,[t^{n/2}/(2\eta)^n]K_n(2\sqrt{t}). \tag{3.13}$$

Thus, the absolute error in taking the first $(n-1)$ terms in (2.11) can be made as small as we wish by taking $(2\eta)$ sufficiently large and the magnitude of the error is smaller than the magnitude of the first term neglected.

It should be noted that in computing the constant $\lambda$ in (2.12) one should carry the computation only to as many terms as are employed in (2.11).

**4. Expansions for $L \geq 0$.** Similar expansions may be obtained in the general case for $L \geq 0$. However we must first make the substitution $u = \rho^L y$ in (1.1). The resulting differential equation is then

$$\rho u'' - 2Lu' + (\rho - 2\eta)u = 0 \tag{4.1}$$

In exactly the same manner as before, we may derive the following expressions for $\rho^{L+1}\Phi_L$ and $\Theta_L$, namely

$$\rho^{L+1}\Phi_L(\eta, \rho) = \frac{(2L+1)!}{(2\eta)^{2L+1}} \rho^{-L} \sum_{n=2L+1}^{\infty} b_n t^{n/2} I_n(2\sqrt{t}), \qquad t = 2\eta\rho \tag{4.2}$$

$$\Theta_L(\eta, \rho) = \lambda_L(\eta)\rho^{-L} \sum_{n=2L+1}^{\infty} b_n(-1)^n t^{n/2} K_n(2\sqrt{t}) \tag{4.3}$$

where $\lambda$ is a constant determined from the condition that $\Theta_L(\eta, 0) = 1$,

$$\lambda_L(\eta) \sum_{n=2L+1}^{\infty} (-1)^n (n-1) b_n/2 = 1. \tag{4.4}$$

and the coefficients $b_n(\eta)$ are obtained from the recurrence relations

$$\begin{aligned} &b_{2L+1} = 1, \qquad b_{2L+2} = 0, \\ &4\eta^2(n-2L)b_{n+1} + nb_{n-1} + b_{n-2} = 0, \qquad n > 2L+2. \end{aligned} \tag{4.5}$$

## REFERENCES


[1] Tables of Coulomb Wave Functions, Vol. 1, National Bureau of Standards Applied Mathematics Series 17, United States Government Printing Office, Washington, 1952.

[2] F. L. Yost, J. A. Wheeler and G. Breit, Coulomb Wave Functions in Repulsive Fields. Phys. Rev. **39**, Jan. 15, 1936.

[3] G. Breit and M. H. Hull, Jr., Asymptotic Expansion of the Irregular Coulomb Function. Phys. Rev. Volume **80**, No. 4, Nov. 15, 1950.

[4] G. Breit and M. H. Hull, Jr., Asymptotic Expansion of Irregular Coulomb Functions for Angular Momentum Zero. Phys. Rev. Volume **80**, No. 3, Nov. 1, 1950.

[5] M. Abramowitz, Asymptotic Expansions of Coulomb Wave Functions. Quarterly of Applied Mathematics, Vol. VII, No. 1, April 1949.

[6] J. David Jackson and John M. Blatt. The interpretation of Low Energy Proton-Proton Scattering. Reviews of Modern Physics, Volume **22**, No. 1, Jan. 1950.

[7] A. Grey and G. B. Mathews, Treatise on Bessel Functions, p. 50, MacMillan, London, 1931.



Computation Laboratory
Washington 25, D. C.
October 1952


(Received February 17, 1953)

# ON A PERTURBATION THEORY BASED ON THE METHOD OF CHARACTERISTICS

C. C. Lin

**1. Introduction*.** It is known that a perturbation theory in the physical plane for supersonic flow and wave propagation is not free from difficulties. A method for improvement has been given by Lighthill[1] and Whitham[2] for dealing with outgoing waves. On the other hand, the characteristic parameters are the natural independent variables to be used for hyperbolic differential equations and it may be expected that in a perturbation theory developed in terms of the characteristic form of the differential equations, such difficulties will not occur.

In the method of characteristics, the desired solution functions $u(x, y)$ and $v(x, y)$ are given in the parametric representation. We introduce a system of characteristic parameters $\alpha$ and $\beta$, and express $x$, $y$, as well as $u$, $v$, in terms of them. In a perturbation theory, the solution thus appears in the form

$$\begin{aligned} u &= u^{(0)}(\alpha, \beta) + \epsilon u^{(1)}(\alpha, \beta) + \cdots \\ v &= v^{(0)}(\alpha, \beta) + \epsilon v^{(1)}(\alpha, \beta) + \cdots \\ x &= x^{(0)}(\alpha, \beta) + \epsilon x^{(1)}(\alpha, \beta) + \cdots \\ y &= y^{(0)}(\alpha, \beta) + \epsilon y^{(1)}(\alpha, \beta) + \cdots \end{aligned} \tag{1.1}$$

Now, such parametric representations have two obvious disadvantages, both of which are often emphasized in the hodograph method. In the first place, the boundary conditions are often specified in the physical plane, and becomes somewhat difficult to impose in the parametric form. Secondly, a single-valued solution in the physical plane could become multiple-valued in the parametric representation (See Cragg's[3] discussion of branch lines in the hodograph method). These difficulties might partly account for the fact that the method of characteristics has not been more extensively used for the development of a perturbation theory. However, if the flow is a perturbation from a nearly uniform state, our freedom in the choice of characteristic parameters makes it possible to avoid multiple-valued mappings of the nature of the branch line in the hodograph method. We observe that, in this case, the characteristic curves are nearly straight in the physical plane. Consequently, the mapping between the plane of characteristic parameters and the physical plane can be made an approximately affine transformation. The mapping between the two planes is expected to be one-to-one and can be easily visualized. In the case of shock formation, associated with the convergence of characteristics in the physical plane, the solution is still expected to be single-valued in the plane of characteristic parameters. Here, the mapping is not one-to-one; but the

* The author wishes to thank Professor K. O. Friedrichs for his suggestions in the final revision of this section.

multiple-valued solution in the physical plane is necessary for the correct representation of the physical phenomena.

A systematic development of a perturbation theory along the lines indicated above is not yet available in the literature, although the ideas are implied in the work of Donov[4] and of Meyer[5], and suggested explicitly by Friedrichs[6,7].

Donov treated the supersonic flow past a two-dimensional thin airfoil, and carried out the calculations to the fourth-order. However, the treatment depends much on the specific nature of the problem at hand, and is not formulated into a general scheme. Furthermore, the characteristic parameters chosen are related to the velocity variables, and hence, when other problems are treated, this approach may suffer from the inconvenience discussed on the preceding page. Meyer treated the problem of waves of finite amplitude in ducts. The method, like that of Lighthill and Whitham, is formulated in terms of successive corrections, rather than treating $u$, $v$ and $x$, $y$ as four functions of two parameters. Thus, his second approximation (to which his solution is carried out) actually corresponds to the first approximation in the form (1.1).

Friedrichs used a coordinate system with one physical distance and one characteristic parameter for the development of a perturbation theory for the flow past a two-dimensional thin airfoil. In this scheme, the solution is found to be very simple. In fact, the flow is approximated by a simple wave up to the second approximation. For higher approximations, he suggested the use of characteristic parameters related to certain linear distances in the physical plane[7].

In the present paper, we shall carry out a systematic development of this idea in the formulation (1.1) with emphasis on the proper choice of the parameters $\alpha$ and $\beta$. The method is applied to a few specific problems to show its advantage over the usual perturbation theory in the physical plane even in the *first* approximation. A convergence proof is given for the case of two-dimensional supersonic flow. The problems of propagation of plane, cylindrical and spherical waves are treated by Miss Phyllis A. Fox[8] by similar methods. This work will be reported separately.*

It should be emphasized here that Lighthill does not limit his approach to the hyperbolic system of differential equations. In fact, he gave examples where the coordinate perturbation leads to improvement of the solution of ordinary differential equations and partial differential equations of the elliptic type. In such cases, there is no "natural" coordinate system, and it appears that the method is sometimes not applicable in the usual form.† However, the Lighthill-Whitham idea is extremely important and valuable, since the method removes serious difficulties in a very elegant and simple manner whenever it can be applied.

* During a visit to M.I.T. on December 17, Dr. Keller showed me some of his old notes dealing with development of suitable methods of successive approximation similar to that in Fox's thesis.

† Cf. Fox's thesis [8]. This kind of conclusion is also reached by Mr. Tozer, working under Professor Lighthill.

Some further comments on coordinate perturbation will be found in the last section.

**2. Perturbation theory in characteristic coordinates.** Consider steady supersonic flow of a gas in two-dimensions. Let $\rho$ be the density, and $u$, $v$ be the components of velocity along the direction of coordinate axes $x$, $y$. We shall limit our discussion to the case of isentropic irrotational flow. Then the pressure $p$ and the speed of sound $c$ are functions of $\rho$. This is also true of the speed $q$, the Mach number $M = q/c$ and the Mach angle

$$A = \sin^{-1}(1/M). \tag{2.1}$$

In the following, we shall use the Mach angle $A$ as the principal variable.

The basic equations may be written in the form

$$\begin{aligned} v_x - u_y &= 0 \\ (c^2 - u^2)u_x - uv(u_y + v_x) + (c^2 - v^2)v_y &= 0. \end{aligned} \tag{2.2}$$

For a small deviation from a state of uniform flow, it is customary to consider a perturbation theory of the form

$$\begin{aligned} u &= U + \epsilon u^{(1)}(x, y) + \cdots \\ v &= \epsilon v^{(1)}(x, y) + \cdots \end{aligned} \tag{2.3}$$

Bernoulli's equation would then give a similar perturbation series for $c^2$. The first approximation of such a perturbation procedure gives solutions of the form

$$\begin{aligned} u^{(1)} &= f(y - \lambda x) + g(y + \lambda x), \\ v^{(1)} &= (1/\lambda)\{-f(y - \lambda x) + g(y + \lambda x)\} \end{aligned} \tag{2.4}$$

where $\lambda = 1/\sqrt{M_0^2 - 1}$, and $M_0$ is the Mach number in the undisturbed stream. The lines

$$y \pm \lambda x = \text{const.} \tag{2.5}$$

are the characteristics in the first approximation.

However, these solutions are inadequate for the representation of certain important cases of small deviations from a uniform flow; e.g., the Prandtl-Meyer expansion. In such a case, the characteristics are definitely not parallel. The reflection of a Prandtl-Meyer flow from a solid boundary is another case in point. These problems require a more exact treatment of the characteristics themselves.

In the $(x, y)$ plane, the characteristic directions $C_\pm$ are given by

$$dy/dx = \tan(\theta \pm A), \tag{2.6}$$

where $\theta$ is the direction of flow. These are only approximately satisfied by (2.5), which would be obtained by setting $\theta = 0$ and $A = A_0$ in (2.6). However, this small deviation in the direction would amount to a large deviation in the coordinates $x$, $y$ if the exact and approximate characteristics are respectively

followed over a long distance. While one family of the characteristics in the case of Prandtl-Meyer flow converges to a point at a finite distance, the approximation (2.5) gives parallel characteristics. In the case of reflection mentioned above, *both* families of characteristics are inadequately approximated by (2.5).

Another way of seeing the inadequacy of the usual perturbation theory is to observe that the velocity and density gradients become infinite as the corner is approached.

The above difficulties would disappear if we use the proper set of coordinates; namely, the characteristic system. If $\alpha$ and $\beta$ are a set of characteristic parameters, the basic equations (2.2) may be written as[7]

$$\begin{cases} y_\alpha \cos(\theta + A) = x_\alpha \sin(\theta + A) \\ y_\beta \cos(\theta - A) = x_\beta \sin(\theta - A) \end{cases} \tag{2.7}$$

$$\begin{cases} v_\alpha \sin(\theta - A) = -u_\alpha \cos(\theta - A) \\ v_\beta \sin(\theta + A) = -u_\beta \cos(\theta + A) \end{cases} \tag{2.8}$$

We may now apply a perturbation theory in the form (1.1), with $u$, $v$, $x$, $y$ as functions of the independent variables $\alpha$, $\beta$; $\theta$ and $A$ being functions of $u$ and $v$. These will be developed in detail for specific examples in the next section. We shall first make the following general observations.

(i) *Relation with other methods.* It is well-known that the characteristics are fixed in the hodograph plane in the present case. This means that (2.8) can be integrated. In fact, we obtain

$$\theta + \psi(A) = F(\alpha), \qquad \theta - \psi(A) = G(\beta) \tag{2.9}$$

where

$$\psi(A) = +\int \frac{\sqrt{M^2 - 1}\, dq}{q} \tag{2.10}$$

In the case $p = K\rho^\gamma$, we have

$$\psi(A) = A - \tfrac{1}{2}\pi + (1/\mu)\tan^{-1}(\mu \cot A), \quad \mu^2 = (\gamma - 1)/(\gamma + 1). \tag{2.11}$$

The constant of integration is here chosen so that $\psi(A) = 0$ at $A = 0$, i.e., the sonic point.

With the solutions (2.9), only the pair of equations (2.7) need be integrated. The present method therefore becomes closely related to the hodograph method. However, it is more useful for the following two reasons. First, the basic importance of the characteristics is explicitly recognized, and hence less trouble is expected with discontinuities. Secondly, the characteristic parameters are still at our disposal, and we may choose them so that the mapping between the characteristic plane and the physical plane is simple. This would avoid unnecessary multiple-valued functions in the mapping process, as it often happens in the hodograph method. Indeed, such a choice may be made so that the usual perturbation theory can be used as a guide. We note that the initial approximation to (2.7) is given by

$$y_\alpha^{(1)} = x_\alpha^{(0)} \tan A_0, \qquad y_\beta^{(0)} = -x_\beta^{(0)} \tan A_0; \tag{2.12}$$

and hence

$$y^{(0)} - \lambda x^{(0)} = k(\beta), \qquad y^{(0)} + \lambda x^{(0)} = h(\alpha), \qquad \lambda = \tan A_0. \tag{2.13}$$

Now if we take $h(\alpha) = h\alpha$, $k(\beta) = k\beta$, the coordinates $\alpha$ and $\beta$ form an oblique Cartesian system obtained from $x^{(0)}$ and $y^{(0)}$ by a linear transformation. In fact, they are exactly the ones suggested by (2.5).

In other problems, such as flow with axial symmetry or spherical waves, the hodograph method is not always useful. The present theory, however, remains essentially the same.

(ii) *The simple wave.* The Prandtl-Meyer problem becomes trivial in the present method, but we shall give the solution here in a form convenient for later reference. Consider the special case $F(\alpha) = F_0$, a constant. Then

$$\theta = \tfrac{1}{2}[F_0 + G(\beta)], \qquad \psi(A) = \tfrac{1}{2}[F_0 - G(\beta)], \tag{2.14}$$

and the first equation of (2.7) gives a family of straight-line characteristics $C_+$. If the angle between the $x$-axis and these characteristics is denoted by $\phi$, then $\phi = \theta + A$. Since (2.14) gives $\theta = F_0 - \psi(A)$, we obtain $\phi = F_0 + A - \psi(A)$, or, by (2.11)

$$\begin{aligned} \tan A &= -\mu \cot \mu(\phi - \phi_*), \qquad \phi_* = \psi(A_0) + \pi/2, \\ \theta &= -[\psi(A) - \psi(A_0)], \end{aligned} \tag{2.15}$$

where the constants have been chosen so that $\theta = 0$ when $A = A_0$. Eq. (215) holds for general simple waves; the dependence of the positional coordinates on $\phi$ changes with the particular solution considered. For waves centered at the origin, it is convenient to introduce the polar coordinate system $(r, \phi)$ by

$$x = r \cos \phi, \qquad y = r \sin \phi. \tag{2.16}$$

The streamlines are given by*

$$r = R \cos^{-\mu^{-2}} \mu(\phi - \phi_*) \tag{2.17}$$

and the cross Mach lines are

$$r = R \cos^{-1\mu^{-2}} \mu(\phi - \phi_*) \sin^{-1} \mu(\phi - \phi_*). \tag{2.18}$$

**3. Application to certain problems in steady supersonic flow.** We shall now apply our method to a few specific problems which cannot be conveniently solved by the usual perturbation theory in the physical plane. The problems are as follows:

(1) interaction of two simple waves,
(2) reflection of a simple wave from a solid boundary,
(3) reflection of a simple wave from a free surface,
(4) the wave length of a supersonic jet.

* Courant-Friedrichs,[7] p. 273. The notation here is somewhat different.

Before we go into the details of any of these problems, let us first develop the general perturbation theory up to the first order.

According to (2.10), the direction and magnitude of the flow are given by

$$\theta = \tfrac{1}{2}[F(\alpha) + G(\beta)], \qquad \psi(A) = \tfrac{1}{2}[F(\alpha) - G(\beta)]. \tag{3.1}$$

If we want to apply these to (2.7), we have to evaluate $\tan(\theta \pm A)$. Now suppose $\theta$ and $A$ differ but slightly from $\theta_0$ and $A_0$, then we may develop $\tan(\theta \pm A)$ into power series in $\theta - \theta_0$ and $A - A_0$. To the first order, we have

$$\tan(\theta \pm A) = \pm\{\tan A_0 + \sec^2 A_0 (A - A_0)\} + \sec^2 A_0 (\theta - \theta_0). \tag{3.2}$$

We also have

$$\begin{aligned} \psi(A) - \psi(A_0) &= \psi'(A_0)(A - A_0), \\ \psi'(A_0) &= -(1 - \mu^2)/(\mu^2 + \tan^2 A_0). \end{aligned} \tag{3.3}$$

Thus, combining (3.1), (3.2) and (3.3), we obtain

$$\begin{aligned} \tan(\theta + A) &= \tan A_0 + \tfrac{1}{2}m[F(\alpha) - F(\alpha_0)] + \tfrac{1}{2}n[G(\beta) - G(\beta_0)], \\ \tan(\theta - A) &= -\tan A_0 + \tfrac{1}{2}n[F(\alpha) - F(\alpha_0)] + \tfrac{1}{2}m[G(\beta) - G(\beta_0)] \end{aligned} \tag{3.4}$$

where $\theta = \theta_0$ and $A = A_0$ for $\alpha = \alpha_0$ and $\beta = \beta_0$, and

$$m = \sec^2 A_0 \,[1 + 1/\psi'(A_0)], \qquad n = \sec^2 A_0 \,[1 - 1/\psi'(A_0)]. \tag{3.5}$$

Notice that $\psi'(A_0) < 0$ and therefore $n > 0$. However, $m$ changes sign as $A_0$ passes through $A_c$, which is given by

$$\tan^2 A_c = 1 - 2\mu^2, \tag{3.6}$$

or

$$M_c = \sqrt{2(1 - \mu^2)/(1 - 2\mu^2)} = \sqrt{4/(3 - \gamma)}. \tag{3.7}$$

For $\gamma = 1.4$, $M_c = 1.58$. For this critical Mach number, *the characteristics are straight lines.*

Let us now substitute (3.4) into (2.7). As indicated in the last section ((2.12) and (2.13)), the initial approximation should be taken as

$$y^{(0)} - \lambda x^{(0)} = k\beta, \qquad y^{(0)} + \lambda x^{(0)} = h\alpha \tag{3.8}$$

where $h$ and $k$ are constants. Now certain conditions will be prescribed for $x$ and $y$, since they are two of the dependent variables sought. Thus, (3.8) in effect imposes certain restrictions on the choice of $\alpha$ and $\beta$. In effect, it means that (at least to an initial approximation) $\alpha$ and $\beta$ must be chosen to be some linear distance along a straight line onto which the points on the initial curve are projected.

The particular choice will depend on the problems in question. It is often found convenient to impose the condition

$$y - \lambda x = k\beta, \qquad y + \lambda x = h\alpha. \tag{3.9}$$

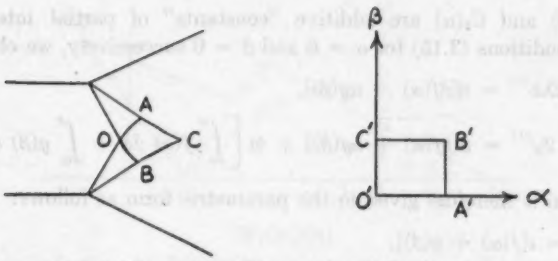



FIG. 1. Interaction of two centered simple waves

Then some of the initial conditions for $x^{(i)}$, $y^{(i)}$ become $x^{(i)} = 0$, $y^{(i)} = 0$ for $i \geqq 1$.

Let us now consider a few specific problems.

(1) *Interaction of two simple waves.* The region of iteration is bounded by four characteristics forming a rectangle in the $(\alpha, \beta)$ plane. Along $OA$ and $OC$, the conditions are specified by the simple wave relations; i.e., $x$, $y$, $A$ and $\theta$ are known as a function of a single parameter along $OA$ and $OC$ (cf. Eqs. (2.15) and (2.18)). We may take this parameter to be $x$ itself and identify it with $\alpha$ along $O'A'$. Then, we have, along $O'A'$,

$$x = \alpha, \qquad \theta = \epsilon f(\alpha), \tag{3.10}$$

while $y$ and $A$ can be calculated from the relations along a characteristic $C_+$. Similarly, along $OC$,

$$x = \beta, \qquad \theta = \epsilon g(\beta). \tag{3.11}$$

We have thus a characteristic initial value problem.

Since $\theta = 0$ at the point 0, $f(0) = g(0) = 0$, and Eq. (3.1) becomes

$$\begin{aligned} \theta &= \epsilon[f(\alpha) + g(\beta)], \\ \psi(A) - \psi(A_0) &= \epsilon[f(\alpha) - g(\beta)]. \end{aligned} \tag{3.12}$$

The initial approximation (3.8) for the coordinates gives

$$x^{(0)} = \alpha + \beta, \qquad y^{(0)} = (\alpha - \beta) \tan A_0 . \tag{3.13}$$

For the next approximation, (2.7) gives

$$\begin{aligned} y_\alpha^{(1)} - \lambda x_\alpha^{(1)} &= mf(\alpha) + ng(\beta), \\ y_\beta^{(1)} + \lambda x_\beta^{(1)} &= nf(\alpha) + mg(\beta). \end{aligned} \tag{3.14}$$

The initial conditions for $x^{(1)}$ are

$$x^{(1)} = 0 \text{ for } \alpha = 0 \text{ and } \beta = 0. \tag{3.15}$$

Now, from (3.14),

$$2\lambda x^{(1)} = n[\beta f(\alpha) - \alpha g(\beta)] + m\left[\int_0^\beta g(\beta)\, d\beta - \int_0^\alpha f(\alpha)\, d\alpha\right] + C_2(\alpha) - C_1(\beta),$$

where $C_1(\beta)$ and $C_2(\alpha)$ are additive "constants" of partial integration. By applying conditions (3.15) for $\alpha = 0$ and $\beta = 0$ successively, we obtain finally

$$(3.16)\qquad \begin{cases} 2\lambda x^{(1)} = n[\beta f(\alpha) - \alpha g(\beta)], \\ 2y^{(1)} = n[\beta f(\alpha) + \alpha g(\beta)] + m\left[\int_0^\alpha f(\alpha)\,d\alpha + \int_0^\beta g(\beta)\,d\beta\right]. \end{cases}$$

The solution is therefore given in the parametric form as follows:

$$(3.17)\qquad \begin{cases} \theta = \epsilon[f(\alpha) + g(\beta)], \\ \psi(A) - \psi(A_0) = \epsilon[f(\alpha) - g(\beta)], \\ x = (\alpha + \beta) + \epsilon(n/2\lambda)[\beta f(\alpha) - \alpha g(\beta)], \\ y = (\alpha - \beta)\lambda + \epsilon(n/2)[\beta f(\alpha) + \alpha g(\beta)] \\ \qquad\qquad + \epsilon(m/2)\left[\int_0^\alpha f(\alpha)\,d\alpha + \int_0^\beta g(\beta)\,d\beta\right], \end{cases}$$

where the functions $f$ and $g$ are given by the conditions

$$(3.18)\qquad \begin{aligned} \theta &= \epsilon f(x) \text{ along the initial characteristic } C_+ \\ \theta &= \epsilon g(x) \text{ along the initial characteristic } C_- \end{aligned}$$

(2) *Reflection of a simple wave from a solid boundary.* This is easily obtained from the last case by taking $g(x) = -f(x)$. Then (3.17) gives $\theta = 0$ along $\alpha = \beta$, which is also the line $y = 0$, by symmetry. Thus, for the case of reflection,

$$(3.19)\qquad \begin{cases} \theta = \epsilon[f(\alpha) - f(\beta)] \\ \psi(A) - \psi(A_0) = \epsilon[f(\alpha) + f(\beta)] \\ x = (\alpha + \beta) + \epsilon(n/2\lambda)[\beta f(\alpha) + \alpha f(\beta)] \\ y = (\alpha - \beta)\lambda + \epsilon(n/2)[\beta f(\alpha) - \alpha f(\beta)] \\ \qquad\qquad + \epsilon(m/2)\left[\int_0^\alpha f(\alpha)\,d\alpha - \int_0^\beta f(\beta)\,d\beta\right]. \end{cases}$$

In particular, along $\alpha = \beta$, $y = 0$, $\theta = 0$, and

$$(3.20)\qquad \begin{aligned} \psi(A) - \psi(A_0) &= 2\epsilon f(\alpha) \\ x &= 2\alpha + \epsilon(n/\lambda)\alpha f(\alpha) \end{aligned}$$

The first formula shows that the pressure deviation is doubled by reflection, in agreement with the usual linear theory. The second formula shows that the width of the interaction region along the solid boundary is *longer* than that indicated by the usual linear theory if the incoming wave is an *expansion*, such as the Prandtl-Meyer flow. It is longer or shorter than that indicated by a geometrical extension of the Prandtl-Meyer fan to the wall depending on whether $M_0$ is larger or smaller than $M_s$. In fact, the following simple geometrical construction can be easily verified.

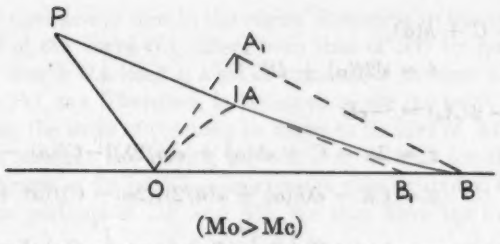



FIG. 2. Reflection of a Prandtl-Meyer expansion from a solid boundary

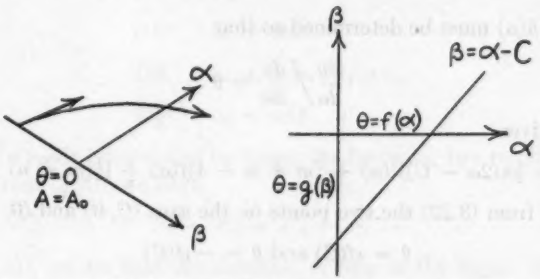



FIG. 3. Reflection from a state of rest

Suppose the Prandtl-Meyer flow is given by the triangle $OPB_1$ in the absence of the wall, with $OA$ as a cross characteristic. The characteristic $PA$ will intersect the wall at $B$ which is obtained by drawing the lines $OA_1$ and $A_1B$ parallel to the exact characteristics at $O$ and $A$ respectively; the point $A_1$ is on a vertical line through $A$. Fig. 2 shows the case $M_0 > M_c$, with the point $A_1$ above $A$. In the case $M_0 = M_c$, the characteristics are straight lines, and the point $B$ coincides with $B_1$.

(3) *Reflection from a region of uniform flow.* This is obtained from (I) by taking

$$g(\beta) = f(\beta + C) + D, \qquad f(C) + D = 0, \tag{3.21}$$

and consider the condition along a curve

$$\beta = \alpha - C + h(\alpha) \tag{3.22}$$

slightly different from the straight line $\beta + C = \alpha$. Then (3.16) give

$$\begin{cases} \theta = \epsilon[f(\alpha) + f(\beta + C) + D] \\ \psi(A) - \psi(A_0) = \epsilon[f(\alpha) - f(\beta + C) - D] \\ x = \alpha + \beta + \epsilon(n/2\lambda)[\beta f(\alpha) - \alpha f(\beta + C) - \alpha D] \\ y = (\alpha - \beta)\lambda + \epsilon(n/2)[\beta f(\alpha) + \alpha f(\beta + C) + 2D] \\ \qquad + \epsilon(m/2)\left[\int_0^\alpha f(\alpha)\, d\alpha + \int_0^\beta f(\beta + C)\, d\beta + D\beta\right]. \end{cases} \tag{3.23}$$

Along $\beta = \alpha - C + h(\alpha)$,

$$
(3.24)\quad \begin{cases} \theta = \epsilon[2f(\alpha) + D] \\ \psi(A) - \psi(A_0) = -\epsilon D \\ x = 2\alpha - C + \epsilon h(\alpha) + \epsilon(n/2\lambda)[-Cf(\alpha) - D\alpha] \\ y = C\lambda - \epsilon\lambda h(\alpha) + \epsilon(n/2)[(2\alpha - C)f(\alpha) + D\alpha] \\ \qquad + \epsilon(m/2)\left[2\int_0^\alpha f(\alpha)\,d\alpha - \int_0^C f(\alpha)\,d\alpha + D(\alpha - C)\right] \end{cases}
$$

The function $h(\alpha)$ must be determined so that

$$\frac{dy}{d\alpha}\Big/\frac{dx}{d\alpha} = \theta$$

In fact, this gives

$$\lambda h'(\alpha) = \tfrac{1}{2}n(2\alpha - C)f'(\alpha) + (m + n - 4)f(\alpha) + [\tfrac{1}{2}(m + n) - 4]D$$

We note that from (3.22) the two points on the axes $(C, 0)$ and $(0, -C)$ give

$$(3.25)\qquad \theta = \epsilon f(C) \text{ and } \theta = -\epsilon f(C)$$

respectively. Such a symmetry relation does not exist in general for other points on the free streamline.

(4) *Wave length of a supersonic jet.* As a final application, we consider a plane supersonic jet when a gas stream is discharged into a gas at rest at a somewhat lower pressure (Fig. 4). If all the characteristics involved were straight, it would be very easy to calculate the length of the jet. Although this is not exactly the case, it is sufficiently accurate for the purpose, if errors of the *second* order are tolerated. This may be seen as follows.

Consider the broken curve *POADEQ*. The portion *PO* is straight and has a slope

$$\lambda_0 = \sqrt{M_0^2 - 1}.$$

The part *OA* is curved, while the part *AD* is straight with the slope

$$\lambda_0 + m\Delta\theta,$$

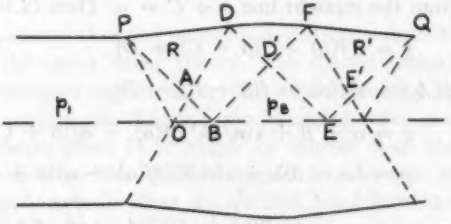



FIG. 4. Supersonic jet

where $\Delta\theta$ is the direction of flow in the region $R$ relative to the original stream. Now, the slope of the curve $OA$ differs from that of $AD$ by quantities of the first order. The length $OA$ itself is a small quantity of the first order compared with the length $PO$, say. Therefore, in trying to locate the point $A$ accurate up to the first order, the slope of $OA$ may be taken to be that of $AD$; i.e., we may regard $OD$ as a straight line having the slope $\lambda_0 + m\Delta\theta$ for the purposes of calculating the length of the jet. The same remark may be applied to the portions $DD'$ and $EE'$ as portions of $DE$ and $EQ$. We then have the following list of slopes:

$$\begin{aligned} PO &\quad \lambda_0 \\ OD &\quad \lambda_0 + m\Delta\theta \\ DE &\quad \lambda_0 + 2\sec^3 A_0(\Delta A) \\ EQ &\quad \lambda_0 - n\Delta\theta \end{aligned}$$

This list can be easily constructed by noting the following two results established previously. From (3.20), we have

$$P_B - P_A = P_A - P_1 ;$$

and from (3.25), we see that the direction of flow in the region $R'$ is obtained from that in the region $R$ by a reflection about a horizontal axis.

We can now easily calculate the distance $PQ$ from simple geometrical considerations. If the semiwidth of the jet is denoted by $W$, and the distance of the point $D$ above the horizontal line $PQ$ is denoted by $h$, (a quantity of the first order), then the length $PQ$ is given by $\Lambda$, where

$$\Lambda - 2\lambda_0 h = \tfrac{1}{2}W[4\lambda_0 + 2\sec^3 A_0(\Delta A) + (m - n)\Delta\theta] = 2[\lambda_0 + \sec^3 A_0(\Delta A)]W$$

Now $h$ is given to the desired degree of approximation by

$$h = (\overline{PD})\cdot\Delta\theta = \lambda_0 W(\Delta\theta).$$

Putting this into the last formula, we find that

$$\Lambda/W = 2[\lambda_0 + \sec^3 A_0(\Delta A)] + 2\lambda_0^2(\Delta\theta),$$

or

$$\Lambda/W = 2\sqrt{\bar{M}^2 - 1}(1 + \sqrt{\bar{M}^2 - 1}\,\Delta\theta), \tag{3.26}$$

where $\bar{M}$ is the Mach number of the flow in the Region $R$.

The formula (3.26) gives the length of the supersonic jet with an error of the second order. Previously, Prandtl gave the formula

$$\Lambda/W = 2\sqrt{M_0^2 - 1} \tag{3.27}$$

and Pack suggested the replacement of $M_0$ by $\bar{M}$. Our formula differs from that of Pack's by an additional correction factor.

Now, Pack used the usual linear theory, and had to be satisfied with an approximation to the jet consisting of sections of trapezoidal form, and he

remarked that "this result does not represent physical reality." Since the difference between Prandtl's formula and Pack's formula is of the order $\epsilon$, it is essential that the improved formula be certainly established including terms of the order $\epsilon$. Unless the solution is accurate up to the order of $\epsilon$, there is always the uncertainty of the reference condition up to the same order. Thus, the choice between $M_0$ and $\bar{M}$ cannot be made conclusive without an accurate solution of the field. From the present standpoint, the usual linear theory is equivalent to the use of the solutions for $(A, \theta)$ accurate to the first approximation and the solutions for $(x, y)$ to the zeroth order. Thus, the accuracy of the calculation of a length can hardly be expected to be conclusive up to the first order. In the present treatment, such difficulties do not occur.

One of Prandtl's experiments gives the value $\Lambda/W = 3.8$. This corresponds to a Mach number $M_0 = 1.85$ and a reservoir pressure 7.74 times the external pressure. Taking $\gamma = 1.40$, we find that $\bar{M} = 1.99$ from this pressure ratio. Prandtl's formula gives the theoretical value 3.1 for $\Lambda/W$, Pack's modification gives the value 3.4 while our formula (3.26) gives the value 3.8, in agreement with experiment. (The value of $\Delta\theta$ can be easily obtained from a table for simple waves.)

Another more complicated formula was obtained by Hasimoto[10] by the hodograph method; it also gives the correct value 3.8 in this case.

**4. Solution by perturbation series.** The characteristic initial value problem formulated in §2 concerns a system of differential equations of the type

$$y_\alpha = P(\alpha, \beta, \epsilon)x_\alpha, \qquad y_\beta = Q(\alpha, \beta, \epsilon)x_\beta; \tag{4.1}$$

where $P(\alpha, \beta, 0)$ and $Q(\alpha, \beta, 0)$ are unequal constants. The initial conditions are

$$x = \alpha \quad \text{along } \beta = 0, \qquad x = \beta \quad \text{along } \alpha = 0. \tag{4.2}$$

The variable $y$ is then fully determined by the condition $y = 0$ at $\alpha = \beta = 0$. If $P(\alpha, \beta, \epsilon)$ and $Q(\alpha, \beta, \epsilon)$ are analytic functions of the parameter $\epsilon$, then according to existing theory,[11] one may expect the solutions $x(\alpha, \beta, \epsilon)$ and $y(\alpha, \beta, \epsilon)$ to be also analytic functions of $\epsilon$, expressible as power series in $\epsilon$, which are convergent for a finite region in the $(\alpha, \beta)$ plane and sufficiently small values of $\epsilon$. Actually, however, it can be shown that the convergence of such series holds for all positive values of $\alpha$ and $\beta$, provided the power series

$$\begin{aligned} P(\alpha, \beta, \epsilon) &= P^{(0)}(\alpha, \beta) + \epsilon P^{(1)}(\alpha, \beta) + \cdots, \\ Q(\alpha, \beta, \epsilon) &= Q^{(0)}(\alpha, \beta) + \epsilon Q^{(1)}(\alpha, \beta) + \cdots, \end{aligned} \tag{4.3}$$

and

$$\begin{aligned} P_\alpha(\alpha, \beta, \epsilon) &= P_\alpha^{(0)}(\alpha, \beta) + \epsilon P_\alpha^{(1)}(\alpha, \beta) + \cdots, \\ Q_\beta(\alpha, \beta, \epsilon) &= Q_\beta^{(0)}(\alpha, \beta) + \epsilon Q_\beta^{(1)}(\alpha, \beta) + \cdots, \end{aligned} \tag{4.4}$$

are convergent for all positive values of $\alpha$ and $\beta$. The usual proof can be extended

to cover this case, but a simple direct proof is given below, since it also gives an indication of the order of magnitude of the successive terms in the solution series. Before proving this, let us first calculate the successive approximations in

$$\begin{aligned} x &= x^{(0)}(\alpha, \beta) + \epsilon x^{(1)}(\alpha, \beta) + \cdots, \\ y &= y^{(0)}(\alpha, \beta) + \epsilon y^{(1)}(\alpha, \beta) + \cdots \end{aligned} \tag{4.5}$$

explicitly.

For the initial approximation, we have

$$y_\alpha^{(0)} = P^{(0)} x_\alpha^{(0)}, \qquad y_\beta^{(0)} = Q^{(0)} x_\beta^{(0)}, \tag{4.6}$$

and the conditions

$$x^{(0)} = \alpha \text{ for } \beta = 0, \qquad x^{(0)} = \beta \text{ for } \alpha = 0. \tag{4.7}$$

The solution is easily verified to be

$$x^{(0)} = \alpha + \beta, \qquad y^{(0)} = P^{(0)}\alpha + Q^{(0)}\beta \tag{4.8}$$

For the higher approximations, $n \geqq 1$, we have

$$\begin{aligned} y_\alpha^{(n)} - P^{(0)} x^{(n)} &= \textstyle\sum_{k=1}^{n} P^{(k)}(\alpha, \beta) x_\beta^{(n-k)}, \\ y_\beta^{(n)} - Q^{(0)} x_\beta^{(n)} &= \textstyle\sum_{k=1}^{n} Q^{(k)}(\alpha, \beta) x_\beta^{(n-k)}, \end{aligned}$$

and the conditions

$$x^{(n)} = 0 \qquad \text{for } \alpha = 0 \text{ and } \beta = 0 \tag{4.10}$$

These equations may be integrated as follows:

$$\begin{cases} y^{(n)} - P^{(0)} x^{(n)} = \displaystyle\sum_{k=1}^{n} \int_0^\alpha P^{(k)} x_\alpha^{(n-k)}\, d\alpha + F^{(n)}(\beta), \\ y^{(n)} - Q^{(0)} x^{(n)} = \displaystyle\sum_{k=1}^{n} \int_0^\beta Q^{(k)} x_\beta^{(n-k)}\, d\beta + G^{(n)}(\alpha); \end{cases} \tag{4.11}$$

which gives

$$\begin{aligned} x^{(n)} = \frac{1}{Q^{(0)} - P^{(0)}} \sum_{k=1}^{n} \left\{ \int_0^\alpha P^{(k)} x_\alpha^{(n-k)}\, d\alpha - \int_0^\beta Q^{(k)} x_\beta^{(n-k)}\, d\beta \right\} \\ + \frac{1}{Q^{(0)} - P^{(0)}} \{F^{(n)}(\beta) - G^{(n)}(\alpha)\}. \end{aligned} \tag{4.12}$$

From the initial conditions (4.10), we see that $F^{(n)}(0) = G^{(n)}(0) = 0$. Putting $\alpha = 0$ and $\beta = 0$ in turn into (4.12), we then obtain

$$F^{(n)}(\beta) = \int_0^\beta Q^{(n)}(0, \beta)\, d\beta, \qquad G^{(n)}(\alpha) = \int_0^\alpha P^{(n)}(0, \alpha)\, d\alpha. \tag{4.13}$$

Equations (4.12) and (4.13) give the explicit solutions for $x^{(n)}(\alpha, \beta)$. From

(4.11), we see that the values of $y^{(n)}$ along the axes are given by (4.13). The solution $y^{(n)}$ is in general given by

$$y^{(n)} = \frac{1}{Q^{(0)} - P^{(0)}} \sum_{k=1}^{n} \left\{ Q^{(0)} \int_0^\alpha P^{(k)} x_\alpha^{(n-k)}\, d\alpha - P^{(0)} \int_0^\beta Q^{(k)} x_\beta^{(n-k)}\, d\beta \right\} + \frac{1}{Q^{(0)} - P^{(0)}} \{ Q^{(0)} F^{(n)}(\beta) - P^{(0)} G^{(n)}(\alpha) \}. \tag{4.14}$$

*Convergence proof.* The convergence proof shall now be carried out for the $x$-series. After this is done, the extension to the $y$-series is obvious.

We first carry out a partial integration for the integrals in (4.12); thus,

$$\int_0^\alpha P^{(k)} x_\alpha^{(n-k)}\, d\alpha = [P^{(k)} x^{(n-k)}]_0^\alpha - \int_0^\alpha P_\alpha^{(k)} x^{(n-k)}\, d\alpha.$$

Since $x^{(n-k)}(0, \beta) = 0$ except when $k = n$, we have

$$\int_0^\alpha P^{(k)} x_\alpha^{(n-k)}\, d\alpha = P^{(k)} x^{(n-k)} - \int_0^\alpha P_\alpha^{(k)} x^{(n-k)}\, d\alpha, \qquad \text{if } k \neq n$$

$$= P^{(n)}(\alpha, \beta) x^{(n-k)} - P^{(n)}(0, \beta)\beta - \int_0^\alpha P_\alpha^{(n)} x^{(0)}\, d\alpha, \quad \text{if } k = n.$$

Thus, for $n \geqq 1$,

$$x^{(n)} = -\frac{1}{Q^{(0)} - P^{(0)}} \sum_{k=1}^{n} \left\{ \int_0^\alpha P_\alpha^{(k)} x^{(n-k)}\, d\alpha - \int_0^\beta Q_\beta^{(k)} x^{(n-k)}\, d\beta \right\} + \frac{1}{Q^{(0)} - P^{(0)}} \{ [F^{(n)}(\beta) - \beta P^{(n)}(0, \beta)] - [G^{(n)}(\alpha) - \alpha Q^{(n)}\alpha, 0)] \}; \tag{4.15}$$

from which we derive

$$|x^{(n)}| \leqq \frac{1}{|Q^{(0)} - P^{(0)}|} \sum_{k=1}^{n} \left\{ \int_0^\alpha |P_\alpha^{(k)}|\,|x^{(n-k)}|\, d\alpha + \int_0^\beta |Q_\beta^{(k)}|\,|x^{(n-k)}|\, d\beta \right\}$$

$$+ \frac{1}{|Q^{(0)} - P^{(0)}|} \left\{ \int_0^\beta |Q^{(n)}(0, \beta)|\, d\beta + \beta\, |P^{(n)}(0, \beta)| + \int_0^\alpha |P^{(n)}(\alpha, 0)|\, d\alpha + \alpha\, |Q^{(n)}(\alpha, 0)| \right\}. \tag{4.16}$$

Consider now the sum

$$S_N = |x^{(0)}| + \sum_{n=1}^{N} \epsilon^n |x^{(n)}|.$$

By using identities of the form

$$\sum_{n=1}^{N} \sum_{k=1}^{n} \epsilon^n a^{(k)} |x^{(n-k)}| = \sum_{k=1}^{N} \sum_{n=k}^{N} \epsilon^n a^{(k)} |x^{(n-k)}| = \sum_{k=1}^{N} \epsilon^k a^{(k)} S_{N-k},$$

We can deduce from (4.8) and (4.13) that

$$
\begin{aligned}
S_N \leqq (\alpha+\beta) &+ \frac{1}{|\,Q^{(0)}-P^{(0)}\,|}\sum_{k=1}^{N} \\
&\cdot\left\{\int_0^{\alpha}\epsilon^k|\,P_\alpha^{(k)}\,|\,S_{N-k}\,d\alpha+\int_0^{\beta}\epsilon^k|\,Q_\beta^{(k)}\,|\,S_{N-k}\,d\beta\right\} \\
&+\frac{\epsilon}{|\,Q^{(0)}-P^{(0)}\,|}\left\{\int_0^{\beta}\left\{\hat{Q}_N(0,\beta)\,d\beta+\beta\hat{P}_N(0,\beta)\right.\right. \\
&\qquad\left.\left.+\int_0^{\alpha}\hat{P}_N(\alpha,0)\,d\alpha+\alpha\hat{Q}_N(\alpha,0)\right\},
\end{aligned}
\tag{4.17}
$$

where

$$
\begin{aligned}
\epsilon\hat{P}_N(\alpha,\beta) &= \epsilon\,|\,P^{(1)}(\alpha,\beta)\,|+\cdots+\epsilon^N\,|\,P^{(N)}(\alpha,\beta)\,|, \\
\epsilon\hat{Q}_N(\alpha,\beta) &= \epsilon\,|\,Q^{(1)}(\alpha,\beta)\,|+\cdots+\epsilon^N\,|\,Q^{(N)}(\alpha,\beta)\,|.
\end{aligned}
\tag{4.18}
$$

From the convergence of the series (4.3), we see that these sums increase to certain limits as $N\to\infty$. Thus, we can find a uniform upper bound to the sums $\hat{P}_N(\alpha,\beta)$ and $\hat{Q}_N(\alpha,\beta)$. This bound may be taken to be independent of $\alpha$, $\beta$ if we are dealing with a finite region in the $\alpha$, $\beta$ plane. Otherwise, we have to introduce the specific conditions that $\lim P_N$ and $\lim Q_N$ are uniformly bounded. These are usually satisfied in practice. Thus if the series (4.3) are bounded as follows

$$
\bar{P}(\alpha,\beta,\epsilon)\leqq|\,P_0\,|+\epsilon J,\qquad \bar{Q}(\alpha,\beta,\epsilon)\leqq|\,Q_0\,|+\epsilon J,
\tag{4.19}
$$

we may write (4.14) as

$$
\begin{aligned}
S_N \leqq &\left(1+\frac{2\epsilon J}{|\,Q^{(0)}-P^{(0)}\,|}\right)(\alpha+\beta) \\
&+\frac{1}{|\,Q^{(0)}-P^{(0)}\,|}\sum_{k=1}^{N}\left\{\int_0^{\alpha}\epsilon^k\,|\,P_\alpha^{(k)}\,|\,S_{N-k}\,d\alpha+\int_0^{\beta}\epsilon^k\,|\,Q_\beta^{(k)}\,|\,S_{N-k}\,d\beta\right\}.
\end{aligned}
\tag{4.20}
$$

We now introduce similar restrictions for the series of the derivative $P_\alpha$ and $Q_\beta$:

$$
\begin{aligned}
[\epsilon\,|\,P_\alpha^{(1)}\,|+\epsilon^2\,|\,P_\alpha^{(2)}\,|+\cdots]<\epsilon J, \\
[\epsilon\,|\,Q_\beta^{(1)}\,|+\epsilon^2\,|\,Q_\beta^{(2)}\,|+\cdots]<\epsilon J.
\end{aligned}
\tag{4.21}
$$

Then, noting that $S_i$ increases monotonically with the index $i$, we may write (3.24) in the form

$$
\begin{aligned}
S_N \leqq &\left(1+\frac{2\epsilon J}{|\,Q^{(0)}-P^{(0)}\,|}\right)(\alpha+\beta) \\
&+\frac{\epsilon J}{|\,Q^{(0)}-P^{(0)}\,|}\left(\int_0^{\alpha}S_N\,d\alpha+\int_0^{\beta}S_N\,d\beta\right).
\end{aligned}
\tag{4.22}
$$

We note that, roughly speaking, the right-hand side of (4.19) differs from $(\alpha + \beta)$ by quantities of the order of $\epsilon$ if we can find a uniform bound for $S_i$. This is the key to the convergence proof, which we carry out by induction.

We try a uniform bound of the form $Ke^{\alpha+\beta}$. This has the advantage of preserving its form upon integration. It is clear that

$$S_0 = \alpha + \beta < Ke^{\alpha+\beta} \tag{4.23}$$

for $K > 1$. Suppose now

$$S_j \leqq Ke^{\alpha+\beta} \text{ for } j = 0, 1, \cdots, N-1. \tag{4.24}$$

Then (4.22) gives

$$S_N \leqq \left(1 + \frac{2\epsilon(J + \bar{J})}{|Q^{(0)} - P^{(0)}|} K\right) e^{\alpha+\beta}$$

This is less than $Ke^{\alpha+\beta}$ if

$$1 + \frac{2\epsilon(J + \bar{J})}{|Q^{(0)} - P^{(0)}|} K < K,$$

or

$$K > \left\{1 - \frac{2\epsilon(J + \bar{J})}{|Q^{(0)} - P^{(0)}|}\right\}^{-1} > 1. \tag{4.25}$$

Thus, if

$$\epsilon < \tfrac{1}{2} |Q^{(0)} - P^{(0)}|/(J + \bar{J}), \tag{4.26}$$

We can indeed find a suitable $K$.

**5. Remarks on coordinate perturbation.** The Lighthill-Whitham approach, with correction made on *one* family of characteristics, leads essentially into the choice of coordinate system used in the theory of Friedrichs. In the present scheme, if the coordinates $(x, y)$ are calculated to the zeroth approximation, we obtain essentially the perturbation theory in the physical plane. If they are calculated to the first approximation, we obtain a solution corresponding to a generalization of the Lighthill-Whitham idea to the correction of both families of characteristics, after the perturbation theory has been developed to the first approximation in the physical plane. (The correction of both families of characteristics was not explicitly mentioned.) When higher approximations are desired, the connection between the Lighthill-Whitham method and a perturbation series is not easily seen, and it is convenient to start with the method of characteristics *ab initio*.

In the case of the Laplace equation, such as the thin airfoil problem considered by Lighthill,[12] the method of coordinate perturbation may be regarded as a perturbation theory in conformal mapping. If $F^{(0)}(z)$ is an analytic function whose imaginary part is the solution for a boundary problem associated with the boundary $C_0$, and $F(z)$ gives the solution for the corresponding problem

with a slightly different boundary $C$, the usual perturbation theory would suggest writing

$$F(z) = F^{(0)}(z) + \epsilon F^{(1)}(z) + \cdots . \tag{5.1}$$

On the other hand, using a conformal mapping between the curves $C_0$ and $C$, we would try to represent the solution as

$$\begin{aligned} F(z) &= F^{(0)}(\zeta) + \cdots, \\ z &= \zeta + \epsilon z^{(1)}(\zeta) + \epsilon^2 z^{(2)}(\zeta) + \cdots . \end{aligned} \tag{5.2}$$

This is essentially the parametric type of solution mentioned in §1.

Such methods can be applied to other problems, e.g., the characteristic value problem for the vibration of a membrane or the Schrödinger equation. Consider, for example, the vibration of a membrane for a nearly circular boundary $C$. Let $C_0$ be the circular boundary in the $\zeta$-plane. The conformal mapping may be written as

$$z = \zeta + \epsilon h(\zeta). \tag{5.3}$$

The problem

$$u_{xx} + u_{yy} + \lambda u = 0, \qquad u = 0 \text{ on } C \tag{5.4}$$

is mapped into

$$u_{\xi\xi} + u_{\eta\eta} + \lambda(1 + \epsilon[h'(\zeta) + \bar{h}'(\zeta)] + \epsilon^2 \mid h'(\zeta) \mid^2)u = 0, \quad u = 0 \text{ on } C_0 . \tag{5.5}$$

Thus, we have the solution in the form

$$\begin{aligned} u &= u^{(0)}(\xi, \eta) + \epsilon u^{(1)}(\xi, \eta) + \epsilon^2 u^{(2)}(\xi, \eta) \\ \lambda &= \lambda^{(0)} + \epsilon\lambda^{(1)} + \cdots \end{aligned} \tag{5.6}$$

The problem now conforms to the usual perturbation theory developed in standard text books.

The method indicated above differs somewhat from the asymptotic process used by Feshbach,[13] which is based on the use of the Green's function for the region in the physical space and is applicable to three-dimensional problems. In fact, the present process is convergent if $\epsilon$ is small enough, and $h(\zeta)$ satisfies certain necessary regularity requirements on the boundary. It is certainly true if $h(\zeta)$ is analytic in a circle slightly larger than $C_0$.

*Note added in proof.* In all the references quoted in §1, the perturbation procedure is based on a small parameter occuring essentially in the *boundary conditions*. After this paper had been sent to the press, Mr. Roger D. Sullivan of the NACA called my attention to his paper *A comparison of two methods of linearized characteristics for a simple unsteady flow*, NACA Technical Note No. 2794 (1952). Although the title would seem to stress the linear approximation, higher approximations are given for a particular problem. Apart from the difference in the physical problems treated, the successive approximations in Mr. Sullivan's paper are based on a parameter in the *differential equations* rather

than in the boundary conditions. Nor is the convergence proof given. But his paper, as well as the present one, is clearly based on the explicit use of the characteristic parameters along the lines suggested by Friedrichs.

## REFERENCES


1. Lighthill, M. J., "A technique for rendering approximate solutions to physical problems uniformly valid", Phil. Mag., ser. 7, vol. **50**, pp. 1179–1201 (1949).
2. Whitham, G. B., "The flow pattern of a supersonic projectile", Comm. Pure and Appl. Math., vol **5**, pp. 301–349 (1952), and the references quoted.
3. Craggs, J. M., "The breakdown of the hodograph transformation for irrotational compressible fluid flow in two dimensions", Proc. Camb. Phil. Soc. **44**, pp. 360–379 (1948).
4. Donov, A., A plane wing with sharp edges in a supersonic stream. Izvestia Akademia Nauk U.S.S.R. Serie Mathematique (1939).
5. Meyer, R. E., "On waves of finite amplitude in ducts", Quart. J. Mech. Appl. Math. **5**, pp. 257–269, 270–291, 1952.
6. Friedrichs, K. O., "Formation and decay of shock waves", Comm. Pure and Appl. Math., vol. **1**, pp. 211–246 (1948).
7. Courant, R. and Friedrichs, K. O., "Supersonic Flow and Wave Propagation", Interscience, New York, 1948, esp. p. 361.
8. Fox, P. A., "On the use of coordinate perturbations in the solution of physical problems", doctoral thesis, M.I.T., November 13, (1953).
9. Pack, D. C. "A note on Prandtl's formula for the wave length of a supersonic jet", Quart. Journ. Mech. and Applied Math., vol. **3**, pp. 173–181, (1950).
10. Hasimoto, Z., "Plane jet of gas emitted from a nozzle with supersonic velocity", Journ. Phys. Soc. Japan, vol. **8**, pp. 394–399, (1953).
11. See Courant, R. and Hilbert, D., "Methoden der Mathematischen Physik" (Interscience, 1937), Chapter V, §§4–7, esp. p. 322, and the references to Lewy and Friedrichs on p. 326.
12. Lighthill, M. J., "A new approach to thin airfoil theory", The Aeronautical Quarterly, vol. **3**, 193–210 (1951).
13. See Morse, P. and Feshbach, H., "Methods of Theoretical Physics" (McGraw-Hill, 1953), pp. 1052–1064; also Chapter IX for bibliography and other related discussions.


Massachusetts Institute of Technology

(Received Dec. 10, 1953)

# LINEARIZATION OF THE EQUATIONS OF NON-STEADY FLOW IN A COMPRESSIBLE FLUID

By John W. Miles

**1. Introduction.** The conditions for linearization of the equations governing the non-steady motion of a thin, two dimensional body in a compressible fluid have been set forth by Lin, Reissner and Tsien.[1] The purpose of the present paper is to extend this analysis to three dimensional planar bodies and to discuss the boundary condition at infinity. (The non-steady motion of slender bodies, such as, *e.g.*, a slender body of revolution has been discussed in an earlier paper,[2] but the present analysis is more complete.)

It is found that sufficient conditions for linearization in the neighborhood of the body are that all of $\delta$, $k\delta$, $M\delta$ and $kM\delta$ be small compared with unity and that any one of $|1 - M|$, $k$ or $\mathcal{A}^{-2}$ be large compared with $\delta^{2/3}$, where $\delta$, $k$, $M$ and $\mathcal{A}$ represent thickness ratio, reduced frequency, free stream Mach number and aspect ratio, respectively. However, $\mathcal{A}^{-1} \gg \delta^{1/3}$ is not a sufficient condition for linearization at large distance from the body.

**2. Formulation.** We assume a continuous, frictionless, non heat-conducting medium in which there are present neither shock waves of finite strength nor body forces. Consequently, the stagnation temperature is constant, the flow is isentropic, and we may assume the existence of a velocity potential ($\phi$). Posing the equation of continuity, the Euler equation, and the isentropic equation of state, it is found that the velocity potential must satisfy the non-linear wave equation

$$a^2\nabla^2\phi = \phi_{tt} + [(\partial/\partial t) + \tfrac{1}{2}(\nabla\phi\cdot\nabla)]\,(\nabla\phi)^2 \tag{2.1},$$

while Bernoulli's equation reads

$$a_0^2 - a^2 = (\gamma - 1)[\phi_t + \tfrac{1}{2}(\nabla\phi)^2] = (\gamma p_0/\rho_0)[1 - (p/p_0)^{(\gamma-1)/\gamma}] \tag{2.2},$$

where $a$ is the sonic velocity, $p$ the pressure, $\rho$ the density, $\gamma$ the specific heat ratio, and the subscript zero designates the stagnation values.

Eqs. (1) and (2) are independent of any particular coordinate system by virtue of the vector formulation, but it is perhaps not out of place to state explicitly that $t$ represents true time, and partial differentiation with respect to $t$ implies that the space coordinates are held fixed during the operation. Now suppose that $(x, y, z)$ is a set of Cartesian coordinates moving with the flight velocity $U$ of the body to be considered. Further, let $l$ be a characteristic length of the order of magnitude of the dimension of this body in the direction

[1] C. C. Lin, E. Reissner and H. S. Tsien, J. Math. Phys. **27**, 220 (1948).
[2] J. W. Miles, J. Aero. Sci. **19**, 280 (1952).

of flight. Then (*cf.* ref. 1), we introduce the dimensionless variables $(\xi, \eta, \zeta, \tau)$ defined by

$$\xi = x/l, \qquad \eta = \lambda y/l, \qquad \zeta = \mu z/l, \qquad \tau = k(Ut/l) \tag{2.3}$$

and a dimensionless perturbation potential $f$ such that

$$\phi(x, y, z, t) = Ux + \epsilon Ulf(\xi, \eta, \zeta, \tau) \tag{2.4}$$

The dimensionless parameters $\lambda$, $\mu$, $k$ and $\epsilon$ are to be defined in a given neighborhood in such a way that $f$ and all of its derivatives with respect to any of $\xi$, $\eta$, $\zeta$ or $\tau$ will be 0(1). (We remark that if such a set of definitions proves impossible, so also will a uniformly valid, small perturbation approximation. However, it may be possible to render the approximation uniformly valid by "straining" the coordinates.[3]) The parameter $\lambda$ will be taken as the reciprocal of the aspect ratio ($AR$) for a wing, while $k$ is a measure of time rate of change and may be chosen as the reduced frequency for harmonic motion. The parameters $\mu$ and $\epsilon$ remain to be determined.

In terms of the foregoing, dimensionless variables, the total vector velocity is given by

$$\mathbf{q} = \nabla\phi = U[\mathbf{i}(1 + \epsilon f_\xi) + \mathbf{j}\epsilon\lambda f_\eta + \mathbf{k}\epsilon\mu f_\zeta] \tag{2.5}$$

The requirement that the velocity perturbation be small (compared with $U$) imposes the initial restrictions

$$\epsilon \ll 1, \qquad \epsilon\lambda \ll 1, \qquad \epsilon\mu \ll 1 \tag{2.6}$$

Moreover, the requirements that both the difference $(a - a_0)$ and the pressure perturbation, as measured by the pressure coefficient

$$c_p = \frac{p - p_0}{\frac{1}{2}\rho U^2} = \left(\frac{2}{\gamma M^2}\right)\left[\left(\frac{a}{a_0}\right)^{2\gamma/(\gamma-1)} - 1\right] \tag{2.7}$$

be small, imposes the additional restrictions

$$\epsilon M^2 \ll 1, \qquad \epsilon k M^2 \ll 1, \qquad \epsilon\lambda M \ll 1, \qquad \epsilon\mu M \ll 1 \tag{2.8}$$

If we now substitute (3) and (4) in (1) and (2), impose the restrictions (6) and (8) and drop all higher order terms, we obtain

$$(\lambda/M)^2 f_{\eta\eta} + (\mu/M)^2 f_{\zeta\zeta} = k^2 f_{\tau\tau} + 2k f_{\xi\tau} + [(1 - M^{-2}) + \epsilon(\gamma + 1) f_\xi] f_{\xi\xi} \tag{2.9}$$

where, for any $\lambda$, $\mu$ and $\epsilon$ satisfying (6) and (8) each term that has been neglected in (9) is definitely small compared with some term that has been retained. However, a term like $\epsilon(\gamma + 1) f_\xi f_{\xi\xi}$ may not be small compared with the other terms in (9), depending on the relative magnitudes of $(1 - M^{-2})$, $(\lambda/M)$, $(\mu/M)$ and $k$.

The corresponding approximation to the pressure coefficient is found to be

$$c_p = -2\epsilon[f_\xi + k f_\tau + \tfrac{1}{2}\epsilon(\lambda^2 f_\eta^2 + \mu^2 f_\zeta^2)] \tag{2.10}$$

The last two terms in (10), albeit second order in velocity may be neglected only

[3] M. J. Lighthill, Phil. Mag. (7) **40**, 1179 (1949); also Aero. Q. **3**, 193 (1951).

after it is established that $\epsilon\lambda^2$ and $\epsilon\mu^2$ are small, as will indeed be the case for every *planar* problem for which a small perturbation solution is possible. (The argument supporting $\epsilon\lambda^2 \ll 1$ follows (3.5), while $\epsilon\mu^2 \ll 1$ is established when $\epsilon$ and $\mu$ are determined in §4.)

**3. The boundary conditions.** We now consider the boundary conditions and the further simplifications that can be effected when the disturbing agent is a nearly plane body, such as a thin wing. Let $\delta$ be a dimensionless parameter defined as the ratio of the larger of the amplitude of the transverse motion* or the thickness of the body to $l$ and specify the surface of the body by

$$z = h(x, y, t) = \delta l g(\xi, \eta, \tau) \tag{3.1}$$

where $g = 0(1)$ by definition (of $\delta$). Now if the body is to be *planar* in the geometrical sense it is clear that both the chord and span must be large compared with the transverse dimension, whence we have the respective restrictions

$$\delta \ll 1 \qquad \lambda\delta \ll 1 \tag{3.2}$$

It is of considerable importance to observe that the foregoing definition of a planar body is inconsistent insofar as very high convexity is implied at the edges. Thus, if the radius of curvature at a wing edge is $0(\delta l)$ it follows that at least one of the second derivatives of $h$ with respect to $x$ and/or $y$ is $0(1/\delta l)$, whence $g_{\xi\xi}$, $g_{\xi\eta}$ or $g_{\eta\eta}$ may be $0(1/\delta^2)$, $0(1/\lambda\delta^2)$ or $0(1/\lambda^2\delta^2)$, respectively, any of which [*cf.* (2)] would violate the initial assumptions. The end result of this inconsistency is the possible failure of a linearized approximation based on (2) in the neighborhood of the wing edges. In particular, it may be necessary to impose some synthetic boundary condition at the edges in order to determine the appropriate singularity there, the nature of which may be restricted on the basis of additional, physical considerations.[4] Alternatively, Lighthill's extension of Poincaré's method (ref. 3) may be applied, and the singularity removed from the external field.

The exact boundary condition at the body, as dictated by the requirement that the flow be tangential at its surface, is given by

$$(\nabla\phi \cdot \mathbf{n}) = n_t, \qquad z = h(x, y, t) \tag{3.3}$$

where $\mathbf{n}$ is the normal to the surface. Evaluating the direction cosines of $\mathbf{n}$, (3) is equivalent to

$$\phi_z \,|_{z=h} = Dh/Dt = h_t + \phi_x h_x + \phi_y h_y \tag{3.4}$$

or, in the dimensionless notation of (1), (2.3) and (2.4),

$$f_\xi \,|_{\xi=\mu g} = (\delta/\mu\epsilon)[kg_\tau + (1 + \epsilon f_\xi)g_\xi + \epsilon\lambda^2 f_\eta g_\eta] \tag{3.5}$$

Now, by definition, $\epsilon f_\xi$ may be neglected compared with unity. Moreover, $\epsilon\lambda^2$ may be neglected, for if it were either dominant or of the same order as the dominant term in (5), we would have $(\mu/\lambda) = 0(\lambda\delta) \ll 1$, and this would imply

* But *cf.* footnote following (3.6) and also last paragraph in paper.

[4] P. A. Lagerstrom, NACA TN 1685 (1948), section I B.

that the lateral extent of the disturbance due to the wing would be small compared with the vertical extent, a condition clearly incompatible with a planar body.

A further simplification of (5) is to impose it at $\zeta = 0$, and this is permissible if it is assumed that $\mu\delta \ll 1$ and that $f$ is regular in the neighborhood of $\zeta = 0$. The former condition will be satisfied automatically (*v.i.*) for all $\mu$ satisfying the other restrictions demanded in the linearization of the *planar* (but not otherwise) problem, while the latter is implicit in the definition of $f$. Moreover, in consequence of the planar restrictions, $g$ must be only two valued in $\xi$ and $\eta$ (corresponding to a non-reentrant profile) and may be decomposed* into the symmetric and antisymmetric (with respect to $\zeta = 0$) functions

$$g^{(s)} = \pm\tfrac{1}{2}(g^+ - g^-), \qquad \zeta = 0\pm \tag{3.6a}$$

$$g^{(a)} = \tfrac{1}{2}(g^+ + g^-), \qquad \zeta = 0\pm \tag{3.6b}$$

where the $+$ and $-$ superscripts correspond to the upper and lower surfaces of the body. Effecting a similar decomposition of the potential, $f$, we obtain, by virtue of the linearity of the approximate boundary condition in both $f$ and $g$, the reduced conditions

$$f_\zeta^{(s)} \mid_{\zeta=0\pm} = \pm(\delta/\mu\epsilon)[g_\xi^{(s)} + kg_\tau^{(s)}] \tag{3.7a}$$

$$f_\zeta^{(a)} \mid_{\zeta=0\pm} = (\delta/\mu\epsilon)[g_\xi^{(a)} + kg_\tau^{(a)}] \tag{3.7b}$$

It is evident from (7) that the potentials $f^{(s)}$ and $f^{(a)}$ must be symmetric and antisymmetric with respect to $\zeta = 0$ and (as is well known) that the two problems can be separated wherever (2.9) can be linearized. (It will be seen below that the pressure equation is linear in $f$ for all such cases.)

In addition to being distinguished by their symmetries, $f^{(s)}$ and $f^{(a)}$ may be expected to exhibit different singularities at the wing edges. Accordingly, as originally observed by Rayleigh,[5] it generally is not possible to construct a solution of a given boundary value problem *via* differentiation of a known solution of some related problem. (This point has been of considerable importance in the electromagnetic diffraction problem.[6])

In addition to the boundary condition at the body, it is necessary to impose appropriate null conditions at infinity, both in order to render the solution unique and to infer restrictions on the linearization. If the differential equation is elliptic it might be sufficient, albeit not always necessary, to impose both the Sommerfeld radiation ("Ausstrahlungsbedingung") and finiteness ("Endlichkeitsbedingung") conditions[7] in a *fixed* coordinate system. However, this usually

* At the same time, the fineness ratio $\delta$ should be split into two parts, $\delta^{(s)}$ and $\delta^{(a)}$. Thus, if the unsteady flow were entirely antisymmetrical only $\delta^{(a)}$ would be required to satisfy $k\delta \ll 1$ and $kM\delta \ll 1$; conversely, $k \gg \delta^{3/2}$ would not be sufficient to justify the linearization of the equation for $\phi^{(a)}$ in transonic range, although that for $\phi^{(a)}$ could be linearized.

[5] Lord Rayleigh, Phil. Mag. **43**, 259 (1897); *Sci. Papers*, **4**, 28.

[6] Rayleigh, Phil. Mag. **44**, 52 (1897); *Sci. Papers* **4**, 326; C. J. Bouwkamp, Physica **12**, 467 (1946); Philips Res. Rep. **5**, 521 (1950); J. W. Meixner, Ann. Phys. Lpz. (6) **6**, 1 (1949); E. T. Copson, Proc. Roy. Soc. Lond. (A) **202**, 277 (1950).

[7] B. B. Baker and E. T. Copson, *Huygens' Principle*, Oxford U. Press (1950), p. 25.

would be awkward, and it frequently appears simpler to introduce a small dissipation parameter and require the solution (*i.e.*, the velocity and pressure perturbation) to vanish exponentially at infinity,[8] after which this parameter may be made to approach zero in the neighborhood of the body (provided, of course, that the limit exists). Whatever the conditions adopted, it should be emphasized that Sommerfield's radiation condition is by itself not sufficient to insure uniqueness. (See, *e.g.*, a recent paper by Ursell.[9]) If the differential equation is hyperbolic the null condition at infinity may be incorporated in the initial conditions, but it must not be overlooked that the differential equation at infinity may be non-linear (*v.i.*).

The fact that the solution must vanish at infinity implies that $f_{\xi\xi}$ in (2.9) never may be neglected throughout the entire field, since, if it were, the boundary conditions (3.5) would yield a solution linear in $\zeta$. Moreover, at least one term on the right side of (2.9) generally must be retained and be of the same order as the left side at large distances from the body, even though all terms on the right prove to be small in the neighborhood of the body. This is evident from the fact that if all terms on the right side of (2.9) were to be neglected Laplace's equation (in $\eta$ and $\zeta$) would result, the solution at infinity then would be governed by the net "source strength" of the body, and $f_\xi^{(o)}$ would become infinite as $\log (\eta^2 + \zeta^2)^{\frac{1}{2}}$ if $\lambda \neq 0$ or as $\zeta$ if $\lambda = 0$. On the other hand, if $\lambda \neq 0$ and only the antisymmetric problem is considered, $f^{(a)}$ is governed asymptotically by the "dipole strength" (falling off inversely with the radius), and it then is permissible to neglect the right side of (2.9) everywhere provided that such neglect is compatible with the boundary conditions at the body.

The preceding discussion implies that the final determination of the parameters $\mu$ and $\epsilon$ depends on the neighborhood in which the solution is examined. Thus, at a sufficiently large distance from the body $\mu$ and $\lambda$ must be equal (The two dimensional problem is of course an exception.), whereas in the neighborhood of the body such equality need not obtain. In the present analysis, we specify $\mu$ and $\lambda$ in the neighborhood of the body, since it is there that the solution usually is required. The situation at large distances from the body then requires to be investigated further only in the event that it proves impossible to satisfy the appropriate null condition at infinity [see (5.5) *et seq.*].

**4. The various special cases.** The value of $\mu$ is now determined by the consideration that both sides of (3.7) must be 0(1). Following ref. 1, we distinguish between two cases, as follows:*

$$A: \quad k = 0(1): \quad \mu = \delta/\epsilon; \quad f_\xi \,|_{\zeta=0+} = g_\xi + k g_\tau \tag{4.1a}$$

$$B: \quad k \gg 1: \quad \mu = k\delta/\epsilon; \quad f_\xi \,|_{\zeta=0+} = g_\tau \tag{4.1b}$$

[8] *Ibid*, p. 154.

[9] F. Ursell, Proc. Cambr. Phil. Soc. **47**, 347 (1951).

* These are based on the assumption that the unsteady flow is due to transverse motion of the wing. On the other hand, if unsteady flow arises from varying incidence of the free stream, as in gust loading, an effective value of $g_\xi$ would be prescribed, and $g_\tau = 0$. In this case, each of the coefficients $c_3 - c_7$ should be multiplied by $k^2$ if $k \gg 1$.

In those cases where the problem is decomposed into its symmetric and antisymmetric parts the boundary conditions are to be imposed at $\zeta = 0+$, and only the upper half space ($\zeta > 0$) need be considered.

The substitution in (2.9) of the alternative values of $\mu$ given by (1) leads to the respective equations

$$f_{\bar{x}\bar{x}} = c_1 f_{\bar{t}} f_{\bar{t}\bar{t}} + c_2 f_{\bar{t}\bar{t}} + c_3 f_{\bar{t}\bar{r}} + \tfrac{1}{2} k c_3 f_{\bar{r}\bar{r}} - c_4 f_{\bar{r}\bar{r}} \tag{4.2a}$$

$$f_{\bar{t}\bar{t}} = c_5 f_{\bar{r}\bar{r}} - c_6 f_{\bar{t}\bar{t}} - c_7 f_{\bar{r}\bar{r}} \tag{4.2b}$$

$$c_1 = (\gamma + 1) M^2 \delta^{-2} \epsilon^3 \tag{4.3}$$

$$c_2 = (M^2 - 1) \delta^{-2} \epsilon^2 \tag{4.4}$$

$$c_3 = 2kM^2 \delta^{-2} \epsilon^2 \tag{4.5}$$

$$c_4 = \delta^{-2} \epsilon^2 \lambda^2 \tag{4.6}$$

$$c_5 = M^2 \delta^{-2} \epsilon^2 \tag{4.7}$$

$$c_6 = k^{-2} \delta^{-2} \epsilon^2 \tag{4.8}$$

$$c_7 = k^{-2} \lambda^2 \delta^{-2} \epsilon^3 \tag{4.9}$$

(The parameters $c_4 - c_7$ are defined differently from those of ref. 1.) The following order of magnitude combinations then manifest themselves,*

$$A_1: \quad c_1 = 1, \quad c_2 = 0(1), \quad c_3 = 0(1), \quad c_4 = 0(1) \tag{4.10}$$

$$A_2: \quad c_1 \ll 1, \quad |c_2| = 1, \quad c_3 = 0(1), \quad c_4 = 0(1) \tag{4.11}$$

$$A_3: \quad c_1 \ll 1, \quad |c_2| \ll 1, \quad c_3 = 1, \quad c_4 = 0(1) \tag{4.12}$$

$$A_4: \quad c_1 \ll 1, \quad |c_2| \ll 1, \quad c_3 \ll 1, \quad c_4 = 1 \tag{4.13}$$

$$B_1: \quad c_5 = 1, \quad c_6 = 0(1), \quad c_7 = 0(1) \tag{4.14}$$

$$B_2: \quad c_5 \ll 1, \quad c_6 = 1, \quad c_7 = 0(1) \tag{4.15}$$

$$B_3: \quad c_5 \ll 1, \quad c_6 \ll 1, \quad c_7 = 1 \tag{4.16}$$

The remainder of the analysis is entirely analogous to that of ref. 1, the conditions imposed by (10)–(16) in conjunction with the restrictions of (2.6), (2.8), and (3.2) leading to the conditions that ensure small perturbations in each case. Accordingly, we present only the final restrictions and the associated

* Other possible combinations of order of magnitude are sub-cases of $A_1 - B_3$. One that may be particularly mentioned is $c_3 = 0(1)$, $c_6 \ll 1$, $c_7 = 1$, which is essentially a subcase of $B_1$ that extends the slender wing theory of $B_1$ to the less severe restriction $kM\mathcal{R} = 0(1)$, rather than $kM\mathcal{R} \ll 1$, and yields the two dimensional wave equation. Similarly, $c_5 = 1$, $c_6$, $c_7 \ll 1$ extends the theory to $(kM\mathcal{R})^{-1} = 0(1)$ provided that $\delta \ll M$ and $kM \gg 1$.

TABLE I.

| | FLOW REGIME | DIFFERENTIAL EQUATION FOR VELOCITY POTENTIAL | WING B.C. $\phi_z\big\|_{z=0}$ | PRESSURE AT WING $\frac{p-p_0}{\rho_0}$ | PRESSURE AT WING: SIMILARITY EXPRESSION FOR $\left(\frac{p-p_0}{\rho_0 U^2 \delta^2}\right)$ | PARAMETRIC RESTRICTIONS: ASPECT RATIO $AR$ ($AR >> \delta$) | MACH NO. $M$ | REDUCED FREQUENCY $k$ | | FINENESS RATIO $\delta$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $A_1$ | NEW LINEAR TRANSONIC | $[(M^2-1)+(\frac{\gamma+1}{\gamma})M^2\phi_x]\phi_{xx}+2(\frac{M}{a_0})\phi_{xt}-\phi_{yy}-\phi_{zz}=0$ | $U\eta_x$ | $U\phi_x$ | $\frac{\delta^{\frac{1}{3}}}{(\gamma+1)^{\frac{1}{3}}}F\left[(\gamma+1)^{\frac{1}{3}}\delta^{\frac{1}{3}}AR, \frac{(M-1)}{(\gamma+1)^{\frac{1}{3}}\delta^{\frac{2}{3}}}, \frac{k}{(\gamma+1)^{\frac{1}{3}}\delta^{\frac{1}{3}}}\right]$ | $AR^{-1}=O(\delta^{\frac{1}{3}})$ | $\lvert M-1\rvert=O(\delta^{\frac{2}{3}})$ | $k=0\ (\delta^{\frac{1}{3}})$ | | $\delta<<1$ |
| $A_2$ | SUBSONIC-$\phi$ SUPERSONIC | $(M^2-1)\phi_{xx}+2(\frac{M}{a_0})\phi_{xt}+\frac{1}{a_0^2}\phi_{tt}-\phi_{yy}-\phi_{zz}=0$ | $U\eta_x+\eta_t$ | $U\phi_x+\phi_t$ | $\frac{\delta}{\lvert M^2-1\rvert^{\frac{7}{3}}}F\left[\lvert M^2-1\rvert^{\frac{1}{3}}AR, \frac{kM^2}{\lvert M^2-1\rvert}, k\right]$ | $AR^{-1}=O(\lvert M^2-1\rvert^{\frac{1}{3}})$ | $\lvert M-1\rvert>>\delta^{\frac{2}{3}}$ $M<<\delta^{-1}$ | $k=0(\lvert 1-M^{-2}\rvert)$ | $k=0(1)$ | $\delta<<1$ $\delta<<M^{-1}$ |
| $A_3$ | UNSTEADY TRANSONIC | $\frac{2}{a_0}\phi_{xt}+\frac{1}{a_0^2}\phi_{tt}-\phi_{yy}-\phi_{zz}=0$ | | | $\delta F(AR,k)$ | $AR^{-1}=O(k^{\frac{1}{3}})$ | $\lvert M-1\rvert<<k$ | $k>>\delta^{\frac{2}{3}}$ | | $\delta<<1$ $\delta<<k^{\frac{3}{2}}$ |
| $A_4$ | SLENDER WING & QUASI-STEADY TRANSONIC | $\phi_{yy}+\phi_{zz}=0$ | | | $\delta AR\,F(k)$ | $AR<<\delta^{-\frac{1}{3}}$ $AR<<\lvert M^2-1\rvert^{-\frac{1}{2}}$ | $\lvert M^2-1\rvert<<AR^{-2}$ | $k<<(MA)^{-2}$ | | $\delta<<1$ |
| $B_1$ | HIGH-FREQUENCY COMPRESSIBLE | $\frac{1}{a_0^2}\phi_{tt}-\phi_{xx}-\phi_{yy}-\phi_{zz}=0$ | $k_0$ | $\phi_t$ | $k^2\delta F(AR,kM)$ | $AR^{-1}=O(kM)$ | $M^{-1}=O(k)$ $M<<(k\delta)^{-1}$ | $k<<(M\delta)^{-1}$ | $1<<k<<\delta^{-1}$ | $\delta<<(kM)^{-1}$ $\delta<<k^{-1}$ |
| $B_1'$ | HIGH-FREQUENCY SLENDER WING | $\frac{1}{a_0^2}\phi_{tt}-\phi_{yy}-\phi_{zz}=0$ | | | $k^2\delta AR\,F(kMAR)$ | $AR<<1$ | $M=O(k^{-1}AR^{-1})$ | | | $\delta<<k^{-1}$ |
| $B_1''$ | HIGH-FREQUENCY QUASI-SLENDER WING | | | | | $AR^{-1}=O(kM)$ | $M>>k^{-1}$ | $k>>M^{-1}$ | | |
| $B_2'$ | HIGH-FREQUENCY INCOMPRESSIBLE | $\phi_{xx}+\phi_{yy}+\phi_{zz}=0$ | | | $k^2\delta F(AR)$ | $AR^{-1}=O(1)$ | $M<<k^{-1}$ | $k<<M^{-1}$ | | |
| $B_2$ | HIGH-FREQUENCY VERY SLENDER WING | $\phi_{yy}+\phi_{zz}=0$ | | | $\text{CONSTANT}\cdot k^2\delta AR$ | $AR<<1$ | $M<<(kR)^{-1}$ | $k<<(MR)^{-1}$ | | |

equations, as set forth in Table 1 (*cf.* Table 1 of ref. 1). The dimensionless potential equations are as follows:

$$A_1: \quad \left\{\frac{2(M-1)}{[(\gamma+1)\delta]^{2/3}} + f_\xi\right\} f_{\xi\xi} + \frac{2k}{[(\gamma+1)\delta]^{2/3}} f_{\xi\tau} - \frac{\lambda^2}{[(\gamma+1)\delta]^{2/3}} f_{\nu\nu} - f_{\xi\xi} = 0 \tag{4.17}$$

$$A_2: \quad \left(\frac{M-1}{|M-1|}\right) f_{\xi\xi} + \frac{2kM^2}{|M^2-1|} f_{\xi\tau} + \frac{k^2M^2}{|M^2-1|} f_{\tau\tau} - \frac{\lambda^2}{|M^2-1|} f_{\nu\nu} - f_{\xi\xi} = 0 \tag{4.18}$$

$$A_3: \quad 2f_{\xi\tau} + kf_{\tau\tau} - (\lambda^2/k) f_{\nu\nu} - 2f_{\xi\xi} = 0 \tag{4.19}$$

$$A_4: \quad f_{\nu\nu} + f_{\xi\xi} = 0 \tag{4.20}$$

$$B_1: \quad f_{\xi\xi} + \lambda^2 f_{\nu\nu} + (kM)^2 (f_{\xi\xi} - f_{\tau\tau}) = 0 \tag{4.21}$$

$$B_2: \quad f_{\xi\xi} + \lambda^2 f_{\nu\nu} + f_{\xi\xi} = 0 \tag{4.22}$$

$$B_3: \quad f_{\nu\nu} + f_{\xi\xi} = 0 \tag{4.23}$$

With a few, minor exceptions, the results given in (4.17)–(4.23) and Table I differ from those of ref. 1 only insofar as the aspect ratio enters. We remark that $k$ enters $A_4$ as a similarity parameter through the boundary conditions.

**5. Conclusions.** Summing up, *sufficient* conditions for linearization of the potential equation *in the neighborhood of a planar body* are

$$\delta \ll 1, \qquad k\delta \ll 1, \qquad M\delta \ll 1, \qquad kM\delta \ll 1 \tag{5.1}$$

and any *one of*[10]

$$|M-1| \gg \delta^{2/3} \tag{5.2a}$$

$$k \gg \delta^{2/3} \tag{5.2b}$$

$$R^{-1} \gg \delta^{1/3} \tag{5.2c}$$

Assuming the satisfaction of these sufficient conditions, the linearized potential equation in the original coordinates reduces to

$$a_0^2 \nabla^2 \phi = (\partial/\partial t + U\partial/\partial x)^2 \phi \tag{5.3}$$

which may be further simplified in the various special cases identified in §4. The pressure equation reduces to

$$p - p_0 = -\rho_0 (\partial/\partial t + U\partial/\partial x)\phi \tag{5.4}$$

Granted the conditions set forth in (1), all three of the conditions set forth in (5.2) can be violated only in the transonic regime, where, by definition, $|1 - M| = O(\delta^{2/3})$. In many situations of practical importance either (5.2b) or (5.2c) still may be satisfied near $M = 1$, although we remark that a relatively

[10] These conditions were stated, without proof, by the author in J. Aero. Sci. **20**, 64 (1953).

small aspect ratio is required in order to render the similarity parameter $[(\gamma + 1)\delta]^{1/3}\mathcal{A}$ small. (But $\mathcal{A}$ should be appropriately computed in establishing this measure; thus, for a delta wing, it would be appropriate to define $\mathcal{A}$ as the semi-apex angle, rather than as the square of the maximum span divided by the area.) If (5.2c) is satisfied and $k = 0(1)$ linearized theory in the transonic regime reduces to "slender wing" theory ($A_4$ in §4), and the antisymmetric problem becomes particularly simple, as first discovered by Jones.[11] (It occasionally has been asserted that a sufficient condition for linearization of the steady flow ($k = 0$) lift problem in transonic flow is $(M^2 - 1)^{1/2}\mathcal{A} \ll 1$, but it appears that $\mathcal{A} \ll \delta^{-1/3}$ will be the more critical condition when $|M - 1| \ll 1$.)

That (5.2c) is a sufficient condition (granted $\delta$ and $k\delta \ll 1$) for the linearization of the transonic lift problem is a consequence of the "dipole" ($r^{-1}$) behaviour of the solution at infinity (*v.s.*, §3). On the other hand, even though (5.2c) permits the reduction of the potential equation to Laplace's equation in the neighborhood of a thin, symmetric wing in steady transonic flow, the "source" solution at infinity is governed by the non-linear equation*

$$\phi_{rr} + r^{-1}\phi_r = [(M^2 - 1) + \{(\gamma + 1)/2U\}\phi_x]\phi_{xx} \tag{5.5}$$

It follows that the solution of the slender body drag problem in steady transonic flow requires the integration of (5.5) for large $r$[12], subject to the conditions that both $r\phi_x$ and $r\phi_r$ be bounded at infinity. The resulting integral then can be expanded in the neighborhood of the body, where the dominant term will be of the form $a_0(x) \log r + b_0(x)$, $a_0(x)$ being determined by the boundary condition at the body and $b_0(x)$ being defined in terms of $a_0(x)$. (*Cf.* Ward's analysis of the supersonic, slender body problem.[13])

Finally, we remark that some of the foregoing conditions stated to be sufficient for linearization may not always be necessary. Thus, it was assumed implicitly that the wing displacement ($h$) and its derivatives had uniform (with respect to $k$) lower bounds of $0(\delta)$, where $\delta$ was the *maximum* displacement, but this may not be true in transient motions of practical importance. *E.g.*, if a sharp edged gust is resolved into its Fourier components having amplitudes $\delta(k)$ $k\delta$ is uniformly bounded, and it would be quite misleading to apply the restriction $k\delta \ll 1$ to the product of the maximum values of $\delta$ and $k$ regarded as independent parameters. In general, the validity of any linearized solution may be determined by *á posteriori* substitution in the original non-linear equations (insofar as it is assumed that these have a unique solution), and considerations such as those of the present paper serve only to indicate the expected results of such a test.

UNIVERSITY OF CALIFORNIA
LOS ANGELES

(Received January 19, 1953)

[11] R. T. Jones, NACA TN 1032 (1946).

* If either (5.2a) or (5.2b) is satisfied, in addition to (5.1), the resulting equation at large distances from the body is linear.

[12] This point has been made independently by K. Oswatitsch in a paper presented at Internat'l Congr. Appl. Mech., Istanbul (1952); private communication *via* P. A. Lagerstrom.

[13] G. N. Ward, Q. J. Mech. and Appl. Math. **2**, 75 (1949).

# STRESS AND VELOCITY FIELDS IN SOIL MECHANICS*

By R. T. Shield

**1. Introduction.** The solutions of many two-dimensional problems in soil mechanics are based upon the assumption that the soil is a plastic material in which slip or yielding occurs when the stresses satisfy the Coulomb formula [1]†

$$f = \tfrac{1}{2}(\sigma_x + \sigma_y)\sin\varphi + \{\tfrac{1}{4}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\}^{\frac{1}{2}} - c\cos\varphi = 0, \tag{1}$$

where $c$ is the cohesion and $\varphi$ is the angle of internal friction of the soil. Neglecting the weight of the soil, the stresses also satisfy the equations of equilibrium

$$\frac{\partial\sigma_x}{\partial x} + \frac{\partial\tau_{xy}}{\partial y} = 0, \qquad \frac{\partial\tau_{xy}}{\partial x} + \frac{\partial\sigma_y}{\partial y} = 0. \tag{2}$$

The two characteristic lines of the hyperbolic system of equations (1), (2) are inclined at an angle $\pi/4 + \varphi/2$ to the direction of the algebraically greater principal stress at any point. As in [2] the characteristic lines will be called the first and second failure lines, with the convention that the direction of the first failure line at a point is obtained from the direction of the algebraically greater principal stress by a clockwise rotation of amount $\pi/4 + \varphi/2$. The inclination of the first failure line to the $x$-axis will be denoted by $\theta$.

Denoting the principal stresses by $\sigma_1$ and $\sigma_2(\sigma_1 < \sigma_2)$, it is convenient to introduce the quantity

$$p = \frac{(\sigma_2 - \sigma_1)}{2\sin\varphi} \geqq 0. \tag{3}$$

An arbitrary constant which has the dimensions of stress is sometimes introduced into the denominator of the expression for $p$ in order to make $p$ nondimensional, but this practice will not be followed here. With the yield condition (1), it can be shown [3] that

$$\begin{aligned} \sigma_x &= -p[1 + \sin\varphi\sin(2\theta + \varphi)] + c\cot\varphi, \\ \sigma_y &= -p[1 - \sin\varphi\sin(2\theta + \varphi)] + c\cot\varphi, \\ \tau_{xy} &= p\sin\varphi\cos(2\theta + \varphi), \end{aligned} \tag{4}$$

and the equations of equilibrium (2) can be replaced by the equations

$$\begin{aligned} &\tfrac{1}{2}\cot\varphi\log p + \theta = \text{const. along a first failure line,} \\ &\tfrac{1}{2}\cot\varphi\log p - \theta = \text{const. along a second failure line.} \end{aligned} \tag{5}$$

* The results presented in this paper were obtained in the course of research sponsored by the Office of Naval Research under Contract N7onr-35801 with Brown University.

† Numbers in square brackets refer to the bibliography at the end of the paper.

These equations were obtained by Massau [4] for a cohesionless soil and by Kötter [5] for a cohesive soil.

Drucker and Prager [6] considered a proper generalization of the Coulomb hypothesis (1) and used the concept of plastic potential [7] to obtain a stress-strain law, assuming that the soil is a perfectly plastic body. In the case of plane strain, the stress-strain law corresponding to the yield function (1) is

$$\epsilon_x = \lambda \frac{\partial f}{\partial \sigma_x}, \qquad \epsilon_y = \lambda \frac{\partial f}{\partial \sigma_y}, \qquad \gamma_{xy} = \lambda \frac{\partial f}{\partial \tau_{xy}}, \tag{6}$$

where $\epsilon_x$, $\epsilon_y$, $\gamma_{xy}$ are the plastic strain rates and $\lambda$ is a non-negative factor of proportionality. With equations (4), the relations (6) can be written

$$\begin{aligned} \epsilon_x &= \tfrac{1}{2}\lambda\{\sin\varphi - \sin(2\theta + \varphi)\}, \\ \epsilon_y &= \tfrac{1}{2}\lambda\{\sin\varphi + \sin(2\theta + \varphi)\}, \\ \gamma_{xy} &= \lambda\cos(2\theta + \varphi). \end{aligned} \tag{7}$$

It will be assumed that there is no deformation of the soil until plastic yielding occurs and therefore

$$\frac{\partial u}{\partial x} = \epsilon_x, \qquad \frac{\partial v}{\partial y} = \epsilon_y, \qquad \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = \gamma_{xy},$$

where $u$, $v$ are the components of velocity along the $x$, $y$-axes. The rate of dilation is found to be

$$\epsilon_x + \epsilon_y = \lambda \sin\varphi \geqq 0, \tag{8}$$

so that plastic deformation must be accompanied by an increase in volume if $\varphi \neq 0$.

It was shown in [2] that the relations (7) imply that the rate of extension along the failure lines is zero. Also, the characteristic lines of the velocities coincide with the characteristic lines of the stresses and it is therefore convenient to refer the velocity equations to the characteristic lines. The orthogonal projections of the velocity vector on the directions of the first and second failure lines passing through the point are denoted by $v_1$ and $v_2$. The velocity projections $v_1$, $v_2$ are related to the cartesian components $u$, $v$ of the velocity by the equations

$$\begin{aligned} v_1 &= u\cos\theta + v\sin\theta, \qquad v_2 = -u\sin(\theta + \varphi) + v\cos(\theta + \varphi), \\ u &= \{v_1\cos(\theta + \varphi) - v_2\sin\theta\}\sec\varphi, \\ v &= \{v_1\sin(\theta + \varphi) + v_2\cos\theta\}\sec\varphi. \end{aligned} \tag{9}$$

The condition that the rate of extension along the failure lines is zero is expressed by the equations

$$\begin{aligned} &dv_1 - (v_1\tan\varphi + v_2\sec\varphi)\,d\theta = 0 \text{ along a first failure line,} \\ &dv_2 + (v_1\sec\varphi + v_2\tan\varphi)\,d\theta = 0 \text{ along a second failure line,} \end{aligned} \tag{10}$$

and these equations, together with the velocity boundary conditions and the condition of non-negative dilatation, determine the velocity field when the failure lines are known.

Discontinuities in the velocity field were considered in [2] and it was shown that a line of discontinuity in the velocity field must be a failure line. The change in velocity across the line must be inclined at an angle $\varphi$ to the line of discontinuity, so that a discontinuity in the tangential velocity is accompanied by a separation or a discontinuity in the normal velocity. In terms of the velocity projections $v_1$ and $v_2$, the requirement that the change in velocity is inclined at an angle $\varphi$ to the line implies that either $v_1$ or $v_2$ is continuous across the line, according as the line is a second or first failure line respectively.

**2. Discontinuities in the stresses.** In the theory of a Prandtl-Reuss material, for which $\varphi = 0$, the possibility of lines of discontinuity in the stress components is well known, and discontinuous stress fields have proved useful in many problems. Discontinuous stress fields are of value when limit analysis is used to obtain lower bounds for the collapse values of the surface tractions in a body or assemblage of bodies, even though the stress fields may be without obvious physical significance. It seems worthwhile, therefore, to consider discontinuous plastic stress fields in a soil, not only from the viewpoint of obtaining real siolutions but also with regard to future applications of limit analysis to soil mechanics. A discussion of stress discontinuities in a cohesionless heavy sol has been given previously by Massau [4]. The following discussion applies to a cohesive soil.

In Fig. 1, the line $l$–$l$ represents an element of a line of stress discontinuity separating the plastic stress fields $a$ and $b$. Subscripts $a$ and $b$ will be used to distinguish the values which a quantity assumes on the two sides of the line.

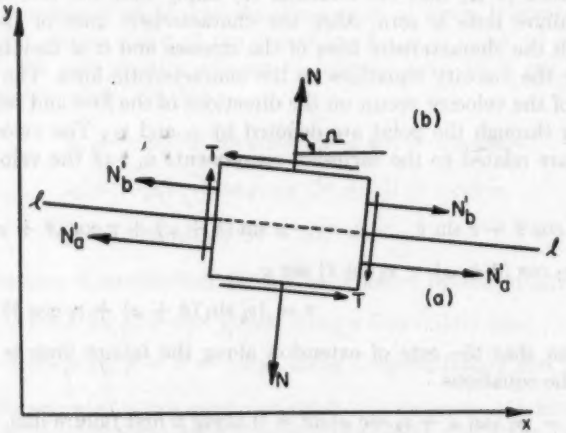



FIG. 1. Stress discontinuity

The normal to the line is inclined at an angle $\Omega$ to the $x$-axis and, from equations (4), the normal and tangential tractions on such a line are

$$N = -p[1 + \sin\varphi \sin(2\theta - 2\Omega + \varphi)] + c\cot\varphi,$$
$$T = p\sin\varphi\cos(2\theta - 2\Omega + \varphi). \tag{11}$$

The equilibrium of the small rectangular element shown in the figure requires that the normal and tangential components of stress $N$ and $T$ are continuous across the line, but the interior components $N'_a$, $N'_b$ may be discontinuous. From equations (11), the equilibrium conditions are

$$p_a[1 + \sin\varphi \sin(2\theta_a - 2\Omega + \varphi)] = p_b[1 + \sin\varphi \sin(2\theta_b - 2\Omega + \varphi)],$$
$$p_a \cos(2\theta_a - 2\Omega + \varphi) = p_b \cos(2\theta_b - 2\Omega + \varphi). \tag{12}$$

The elimination of $p_a$, $p_b$ between these two equations gives, after some reduction,

$$\sin(\theta_a + \theta_b - 2\Omega + \varphi) + \sin\varphi\cos(\theta_a - \theta_b) = 0, \tag{13}$$

provided that $\sin(\theta_a - \theta_b) \neq 0$, that is, provided that the discontinuity is not of zero strength.

The condition (13) together with one of the conditions (12) correspond to the jump conditions established by Prager [8] for a Prandtl-Reuss material ($\varphi = 0$). For zero angle of friction, condition (13) becomes

$$\theta_a + \theta_b = 2\Omega \pm n\pi. \tag{14}$$

The Mohr's circles for the regions $a$ and $b$ are shown in Fig. 2, where $p_a$ is taken greater than $p_b$ for definiteness. The circles touch the yield locus

$$|\tau| = c - \sigma\tan\varphi,$$

since the regions are at the point of yielding. The stress point $C$ has the coordinates $(N, -T)$ and the circles intersect at this point. The poles of the two

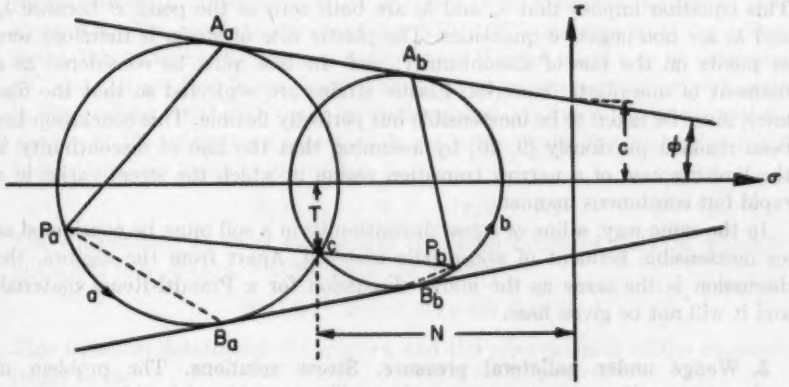



FIG. 2. Mohr's circles for Fig. 1

circles are obtained by drawing a line through $C$ parallel to the element $l$–$l$ of the line of discontinuity. The points $P_a$, $P_b$ where this line meets the circles are the poles of the circles. If $A_a$, $B_a$ and $A_b$, $B_b$ denote the points of contact of the circles $a$ and $b$ with the yield locus, then the lines $P_aA_a$, $P_aB_a$ and $P_bA_b$, $P_bB_b$ give the directions of the first and second failure lines in the regions $a$ and $b$ respectively. It can be seen from the diagram that the direction of the line of discontinuity lies in the acute angle formed by the failure lines in region $a$ and in the obtuse angle formed by the failure lines in region $b$.

It can easily be shown that a failure line cannot be a line of discontinuity in the stresses. Since a discontinuity in the velocity field can occur only across a failure line, it follows that the velocity field must be continuous across a line of stress discontinuity. For a straight line of discontinuity, which may be taken to be parallel to the $x$-axis, the continuity of the velocity component $u$ across the line implies that $\partial u/\partial x = \epsilon_x$ is continuous across the line. This argument can be extended to show that $\epsilon_x$ is continuous across a curved line of stress discontinuity at any point $P$ of the line, where the $x$-axis is taken parallel to the tangent to the line at $P$.

Considering first a Prandtl-Reuss material, equations (7) give

$$\epsilon_x = -\tfrac{1}{2}\lambda \sin 2\theta,$$

since $\varphi$ is zero for this material. It follows that

$$\lambda_a \sin 2\theta_a = \lambda_b \sin 2\theta_b, \tag{15}$$

where $\lambda_a$, $\lambda_b$ and $\theta_a$, $\theta_b$ are the values assumed by $\lambda$ and $\theta$ in the two stress fields at the point $P$. The jump condition (14) on $\theta_a$, $\theta_b$ requires that

$$\theta_b = -\theta_a \pm n\pi. \tag{16}$$

Since the line is not a failure line, $\sin 2\theta_a \neq 0$, and the substitution of (16) in (15) gives

$$\lambda_a = -\lambda_b.$$

This equation implies that $\lambda_a$ and $\lambda_b$ are both zero at the point $P$ because $\lambda_a$ and $\lambda_b$ are non-negative quantities. The plastic rate of strain is therefore zero at points on the line of discontinuity, and the line must be considered as a filament of non-plastic material. Elastic strains are neglected so that the filament must be taken to be inextensible but perfectly flexible. This conclusion has been reached previously [9, 10] by assuming that the line of discontinuity is the limiting case of a narrow transition region in which the stress varies in a rapid but continuous manner.

In the same way, a line of stress discontinuity in a soil must be considered as an inextensible filament of non-plastic material. Apart from the algebra, the discussion is the same as the above discussion for a Prandtl-Reuss material, and it will not be given here.

**3. Wedge under unilateral pressure. Stress solutions.** The problem of wedges with uniform pressure on one face will be considered in this section and

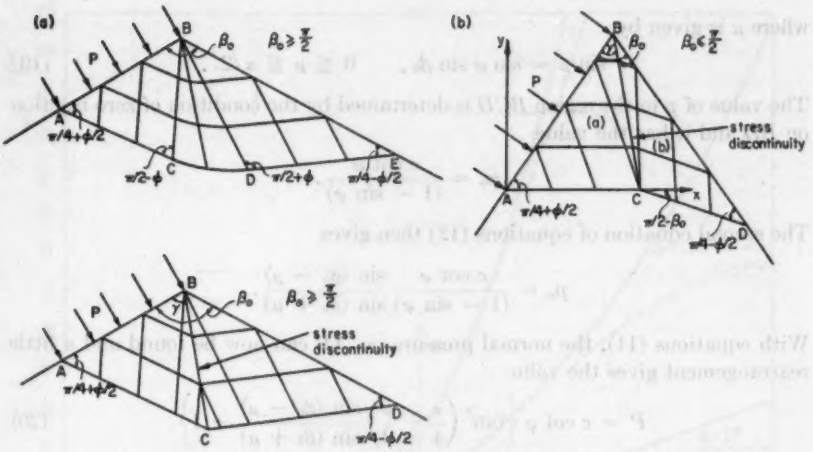



FIG. 3. Stress fields for loaded wedge

in the following section. A continuous stress solution for obtuse angled wedges has been given by Prandtl [11] and the pattern of the failure lines is shown in Fig. 3a. The wedge $ABE$ of angle $\beta_0 (\beta_0 \geqq \pi/2)$ is loaded by a uniform pressure $P$ along $AB$, producing a plastic state of stress in the region $ABEDC$. The regions $ABC$, $BDE$ are regions of constant stress and the region $BCD$ is a zone of radial shear of angle $\beta_0 - \pi/2$. The pressure $P$ required to produce plastic flow with this pattern of failure lines is given by

$$P = c \cot \varphi \{\exp [(2\beta_0 - \pi) \tan \varphi] \tan^2 (\tfrac{1}{4}\pi + \tfrac{1}{2}\varphi) - 1\}. \tag{17}$$

The stress field of Fig. 3a is not applicable to acute angled wedges $(\beta_0 < \pi/2)$ since the two constant stress regions $ABC$, $BDE$ would overlap. However, the introduction of a line of stress discontinuity in the plastic stress region enables a stress solution to be found. Fig. 3b shows the wedge loaded along $AB$. The constant stress regions $ABC$, $BCD$ are separated by a line of stress discontinuity $BC$ which is inclined to $AB$ at an angle $\gamma$ to be determined. $AC$ and $CD$ are first failure lines. The different values a quantity may assume in the regions $ABC$, $BCD$ will be distinguished by the subscripts $a$ and $b$ respectively. Choosing the $x, y$-axes as shown in the figure, the inclination $\theta$ of the first failure lines assumes the values

$$\theta_a = 0, \qquad \theta_b = -(\pi/2 - \beta_0).$$

The substitution of these values into the jump condition (13), in which $\Omega$ is put equal to $\gamma - (\pi/4 - \varphi/2)$, gives the relation

$$\sin (\beta_0 - 2\gamma) + \sin \varphi \sin \beta_0 = 0.$$

This equation determines the angle $\gamma$ and the relevant root of the equation is found to be

$$\gamma = \beta_0/2 + \mu/2, \tag{18}$$

where $\mu$ is given by

$$\sin \mu = \sin \varphi \sin \beta_0, \qquad 0 \leqq \mu \leqq \pi/2. \tag{19}$$

The value of $p$ in the region $BCD$ is determined by the condition of zero traction on $BD$ and it has the value

$$p_b = \frac{c \cot \varphi}{(1 - \sin \varphi)}.$$

The second equation of equations (12) then gives

$$p_a = \frac{c \cot \varphi}{(1 - \sin \varphi)} \frac{\sin (\beta_0 - \mu)}{\sin (\beta_0 + \mu)}.$$

With equations (11), the normal pressure on $AB$ can now be found and a little rearrangement gives the value

$$P = c \cot \varphi \left\{ \tan^2 \left( \frac{\pi}{4} + \frac{\varphi}{2} \right) \frac{\sin (\beta_0 - \mu)}{\sin (\beta_0 + \mu)} - 1 \right\}. \tag{20}$$

When $\varphi = 0$ (Prandtl-Reuss material), $\mu$ is also zero and (20) takes the form

$$P = 2c(1 - \cos \beta_0),$$

agreeing with the value obtained by Prager [8]. For small values of $\beta_0$, expression (20) is approximately

$$P = c\beta_0^2 \cos \varphi,$$

and it follows that for small values of $\beta_0$, the ratio $P/c$ decreases as the angle

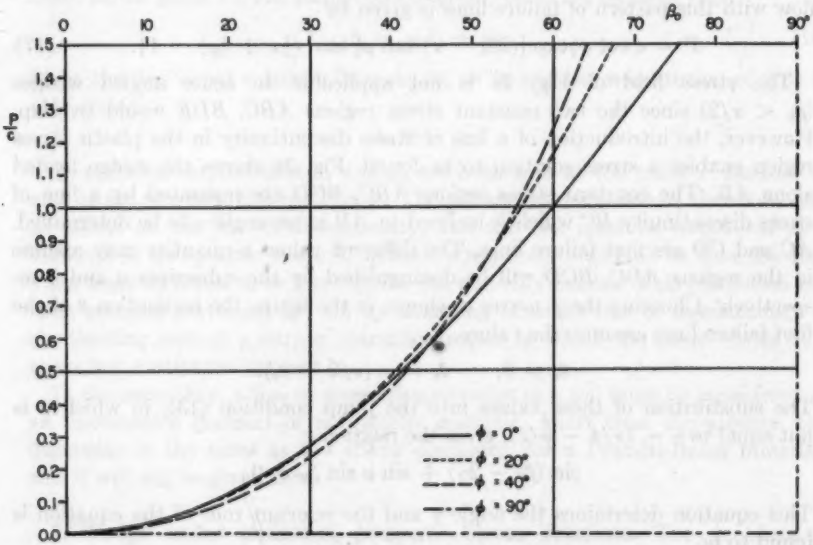



FIG. 4a. Pressure on face of acute angled wedge

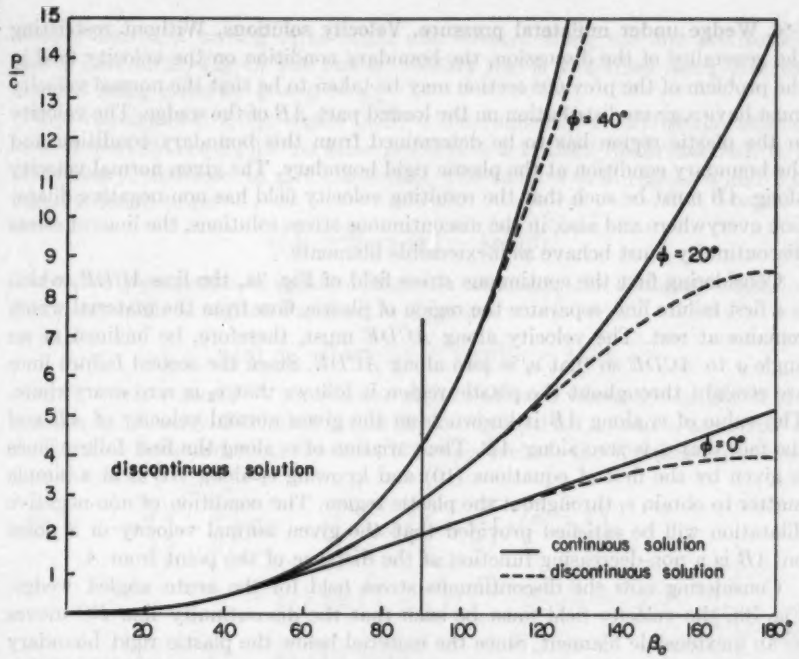



FIG. 4b. Pressure on wedge face as given by the discontinuous and continuous solutions

of friction $\varphi$ is increased. Fig. 4a shows the variation of $P/c$ with the angle of the wedge for $\varphi = 0°$, 20° and 40°. The limiting case $\varphi = 90°$ is also shown in the diagram.

When $\beta_0 = \pi/2$, the expressions (17) and (20) have the common value $2c \tan(\pi/4 + \varphi/2)$, and the discontinuous solution is a continuation of the continuous solution. The full lines in Fig. 4b shows the variation in $P/c$ with the wedge angle $\beta_0$ as furnished by the two solutions for angles of friction of 0°, 20° and 40°. The discontinuous solution also satisfies all the stress conditions when the angle of the wedge is greater than $\pi/2$, and the stress field, Fig. 3c, is an alternative solution to the continuous solution of Fig. 3a. However, it will be seen that a velocity field cannot be associated with the stress field of Fig. 3c, so that the discontinuous solution is not acceptable physically for wedge angles greater than a right angle. The pressure $P$ as furnished by the discontinuous solution for obtuse angled wedges is again given by (20) and this value is lower than the value (17) obtained from the continuous solution. The broken lines in Fig. 4b give the values of $P/c$ for varying $\beta_0 (\beta_0 \geqq \pi/2)$ for angles of friction of 0°, 20° and 40°.

It should perhaps be pointed out that the line of discontinuity $BC$ in Fig. 3c is not the line joining the point $B$ in Fig. 3a to the point of intersection of the lines $AC$, $ED$ except when the angle of friction is zero.

**4. Wedge under unilateral pressure. Velocity solutions.** Without restricting the generality of the discussion, the boundary condition on the velocity field in the problem of the previous section may be taken to be that the normal velocity must have a given distribution on the loaded part $AB$ of the wedge. The velocity in the plastic region has to be determined from this boundary condition and the boundary condition at the plastic rigid boundary. The given normal velocity along $AB$ must be such that the resulting velocity field has non-negative dilatation everywhere and also, in the discontinuous stress solutions, the lines of stress discontinuity must behave as inextensible filaments.

Considering first the continuous stress field of Fig. 3a, the line $ACDE$, which is a first failure line, separates the region of plastic flow from the material which remains at rest. The velocity along $ACDE$ must, therefore, be inclined at an angle $\varphi$ to $ACDE$ so that $v_2$ is zero along $ACDE$. Since the second failure lines are straight throughout the plastic region it follows that $v_2$ is zero everywhere. The value of $v_1$ along $AB$ is known from the given normal velocity of $AB$ and the fact that $v_2$ is zero along $AB$. The variation of $v_1$ along the first failure lines is given by the first of equations (10) and knowing $v_1$ along $AB$ it is a simple matter to obtain $v_1$ throughout the plastic region. The condition of non-negative dilatation will be satisfied provided that the given normal velocity of a point on $AB$ is a non-decreasing function of the distance of the point from $A$.

Considering now the discontinuous stress field for the acute angled wedge, Fig. 3b, the velocity field must be such that the discontinuity line $BC$ moves as an inextensible filament. Since the material below the plastic rigid boundary $ACD$ remains at rest, this implies that the velocity component along $BC$ is zero for points on $BC$. In determining the velocity field it is convenient to use

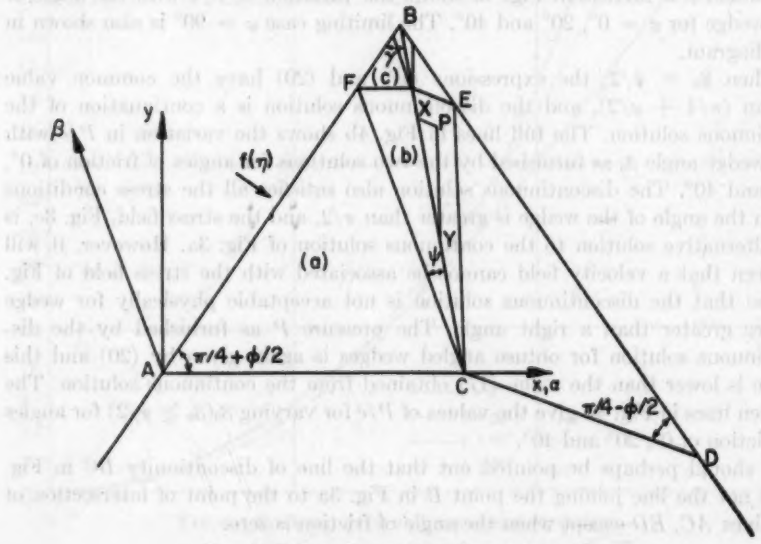



FIG. 5. Failure line pattern for acute wedge

oblique axes $\alpha$, $\beta$ having $A$ as origin and directed along the first and second failure lines at $A$, Fig. 5. The normal velocity $f(\eta)$ is prescribed along $AB$, the line $AB$ being given by $\alpha = \beta = \eta$.

The velocity field in the region $ABC$ is determined in a very similar manner to that used by Lee [9] for the same problem in a Prandtl-Reuss material. It is unnecessary, therefore, to give the details. The value of $v_1$ in regions $a$ and $b$ of Fig. 5 is found to be

$$v_1(\beta) = 2 \cos (\tfrac{1}{4}\pi - \tfrac{1}{2}\varphi) f(\beta).$$

The velocity projection $v_2$ is zero in region $a$ because of the boundary condition along $AC$, while in regions $b$ and $c$ it is given by

$$v_2(\alpha) = -2 \cos (\tfrac{1}{4}\pi - \tfrac{1}{2}\varphi) t f[(\alpha - 1)/t],$$

where $t = \sin \psi \sec (\psi - \varphi)$, $\psi$ is the angle $BCF$ and where the length of $AC$ is taken to be unity. For other points in the region $ABC$, $v_1$ and $v_2$ are given in the form of finite series. As the apex $B$ is approached, the limiting expressions take the form

$$\left.\begin{aligned} v_1(\beta) &= 2 \cos\left(\frac{\pi}{4} - \frac{\varphi}{2}\right) f(\beta) - t\left[2 \cos\left(\frac{\pi}{4} - \frac{\varphi}{2}\right) f\left(\frac{\beta - 1}{t}\right)\right. \\ &\qquad \left. - t\left\{2 \cos\left(\frac{\pi}{4} - \frac{\varphi}{2}\right) f\left(\frac{(\beta - 1)/t - 1}{t}\right) \cdots\right\}\right], \\ v_2(\alpha) &= -t\left[2 \cos\left(\frac{\pi}{4} - \frac{\varphi}{2}\right) f\left(\frac{\alpha - 1}{t}\right)\right. \\ &\qquad \left. - t\left\{2 \cos\left(\frac{\pi}{4} - \frac{\varphi}{2}\right) f\left(\frac{(\alpha - 1)/t - 1}{t}\right) \cdots\right\}\right]. \end{aligned}\right\} \tag{21}$$

The continued fractions which are the arguments of the function $f$ in these expressions represent the coordinates of points on $AB$ after successive reflections at $BC$ and $AB$ along lines parallel to the oblique axes at $A$.

It can readily be shown that the velocities $v_1$, $v_2$ approach limits at the vertex $B$. Also the limits are such that the velocity of the plastic field near $B$ approaches the velocity of the non-plastic filament $BC$ at $B$, this velocity being uniquely determined by the velocity condition on $BC$ and the given normal velocity of $AB$ at $B$.

The velocities of the points on the filament $BC$ are now completely determined. The velocity of a point $P$ in the region $BCD$ is determined from the velocity of the filament $BC$ and the boundary condition at the plastic rigid boundary $CD$. The velocity at $P$ depends only upon the velocities of the points $X$, $Y$ where the failure lines through $P$ meet the line $BC$ (see Fig. 5). $v_1$ is constant along $PX$, $v_2$ is constant along $PY$ and $v_1$ and $v_2$ are known at the points $X$, $Y$. If $P$ lies in the region $CDE$ then $v_2$ is zero.

The function $f(\eta)$ which specifies the distribution of normal velocity along $AB$ can not be chosen in a completely arbitrary manner. Certain restrictions must be placed upon $f(\eta)$ in order that the dilatation be positive everywhere in the plastic region. The restrictions are most easily obtained in terms of the velocity

$\chi(l)$ of the non-plastic filament at the point distance $l$ from $C$. $\chi(l)$ is continuous since the filament is non-plastic and also $\chi(0) = 0$ from the condition at $C$. Omitting the details of the analysis, it is found that the dilatation will be positive everywhere provided that $\chi'(l) \geqq 0$ and $\chi'(l)$ is a non-decreasing function. These conditions imply that the rate of change of curvature of the filament increases towards the vertex. Thus the wedge behaves in a similar fashion to a beam which is bent by uniform loading.

The restrictions which have to be placed upon $f(\eta)$ are obtained by expressing $f(\eta)$ in terms of the restricted function $\chi(l)$.

In the discontinuous stress solution of Fig. 3c, the velocity conditions for the problem of the acute angled wedge apply in this case also. The plastic rigid boundary $AC$ enforces $v_2 = 0$ along $AC$ and it follows that $v_2 = 0$ throughout region $ABC$ of Fig. 3c, since the failure lines are straight. The condition that the velocity component along $BC$ is zero for points on $BC$ then requires $v_1$ to be zero on $BC$ and in consequence $v_1$ must be zero throughout the region $ABC$. Thus the velocity field obtained requires $f(\eta) = 0$ along $AB$ and the solution does not constitute plastic flow. The discontinuous stress solution for the obtuse angled wedge is therefore physically inadmissible.

**5. Loaded trapezoid.** A further example of discontinuous stress fields is shown in Fig. 6a. The trapezoid $ABCD$ is in a plastic state of stress due to the normal pressures $P$, $Q$ on the parallel sides $AB$, $CD$, the sides $AD$, $BC$ being free from applied traction. The lines $AO$, $BO$, $CO$, $DO$ are lines of stress discontinuity

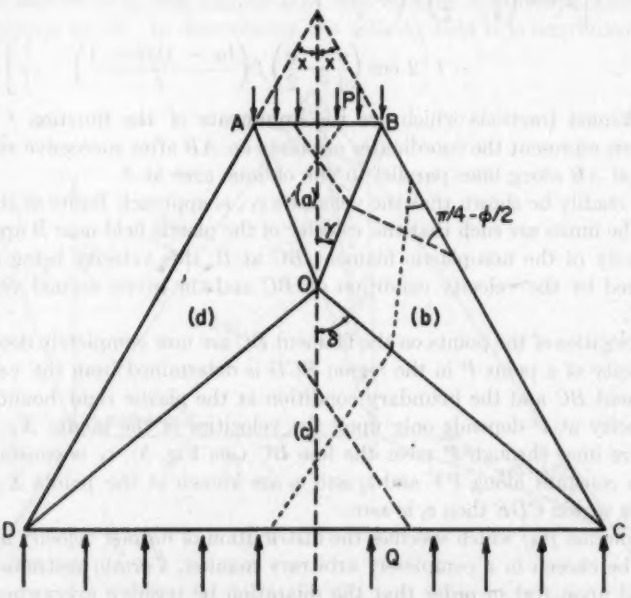



FIG. 6a. Discontinuous stress system for loaded trapezoid

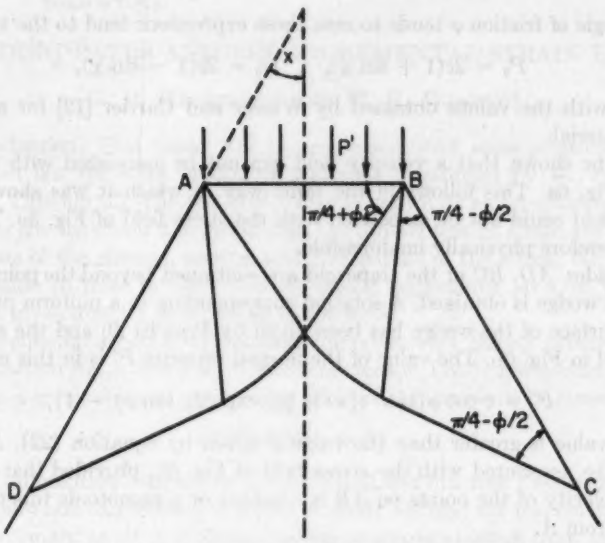



FIG. 6b. Continuous stress field for truncated wedge

separating the regions $a$, $b$, $c$, $d$ of constant stress, where the angles marked $\gamma$, $\delta$ in the figure remain to be determined.

The stress boundary conditions require that

$$\theta_a = -\left(\frac{\pi}{4}+\frac{\varphi}{2}\right), \qquad \theta_b = -\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)+\chi, \qquad \theta_c = -\left(\frac{\pi}{4}+\frac{\varphi}{2}\right),$$

where $2\chi$ is the angle between the sides $AD$, $BC$ and where the subscripts refer to the regions marked $a$, $b$, $c$ in the figure. With these values of $\theta$, the jump condition (13) at the discontinuity lines $BO$, $CO$ show that

$$\cos(2\gamma+\chi) = \sin\varphi\cos\chi,$$
$$\cos(2\delta-\chi) = \sin\varphi\cos\chi,$$

and the relevant roots of these equations are

$$\gamma = \tfrac{1}{2}\nu - \tfrac{1}{2}\chi, \qquad \delta = \tfrac{1}{2}\nu + \tfrac{1}{2}\chi,$$

where

$$\cos\nu = \sin\varphi\cos\chi, \qquad 0 \leqq \nu \leqq \pi/2.$$

The values of $p$ in the regions $a$ and $c$ can be deduced from the known value of $p$ in region $b$ and the jump conditions (12) at the lines $BO$, $CO$. The pressures $P$, $Q$ can then be obtained and it is found that

$$\left.\begin{aligned} P &= c\cot\varphi\left\{\tan^2\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)\frac{\sin(\nu+\chi)}{\sin(\nu-\chi)}-1\right\},\\ Q &= c\cot\varphi\left\{\tan^2\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)\frac{\sin(\nu-\chi)}{\sin(\nu+\chi)}-1\right\}. \end{aligned}\right\} \tag{22}$$

As the angle of friction $\varphi$ tends to zero these expressions tend to the values

$$P_0 = 2c(1 + \sin \chi), \qquad Q_0 = 2c(1 - \sin \chi),$$

agreeing with the values obtained by Winzer and Carrier [12] for a Prandtl-Reuss material.

It can be shown that a velocity field can not be associated with the stress field of Fig. 6a. This follows in the same way in which it was shown that a velocity field could not be associated with the stress field of Fig. 3c. The stress field is therefore physically inadmissible.

If the sides $AD$, $BC$ of the trapezoid are continued beyond the points $D$, $C$ a truncated wedge is obtained. A solution corresponding to a uniform pressure on the top surface of the wedge has been given by Prandtl [9] and the stress field is outlined in Fig. 6b. The value of the normal pressure $P'$ is in this case

$$P' = c \cot \varphi \{\tan^2 (\tfrac{1}{4}\pi + \tfrac{1}{2}\varphi) \exp (2\chi \tan \varphi) - 1\},$$

and this value is greater than the value $P$ given by equation (22). A velocity field can be associated with the stress field of Fig. 6b. provided that the given normal velocity of the points on $AB$ is constant or a monotonic function of the distance from $A$.

## BIBLIOGRAPHY

[1] K. Terzaghi, "Theoretical Soil Mechanics", John Wiley and Sons, p. 22, 1943.
[2] R. T. Shield, "Mixed Boundary Value Problems in Soil Mechanics", Q. Appl. Math., **11**, No. 1 (1953).
[3] V. V. Sokolovsky, "Statics of Earthy Media", Izdatelsvo Akademii Nauk S.S.R., Moscow, 1942.
[4] J. Massau, "Mémoire sur l'intégration graphique des équations aux derivées partielles", Edition du Centenaire, Comité National de Mécanique, Bruxelles, pp. 121–391, 1952 (reprinted from Ann. de l'Ass. des Ing. sortis des Ecoles spéc. de Gand, 1900–1904).
[5] F. Kötter, "Die Bestimmung des Drucks an gekruemmten Gleitflaechen, eine Aufgabe aus der Lehre vom Erddruck", Berlin Akad. Berichte, 229–233 (1903).
[6] D. C. Drucker and W. Prager, "Soil Mechanics and Plastic Analysis or Limit Design", Q. Appl. Math., **10**, pp. 157–165 (1952).
[7] R. von Mises, "Mechanik der plastischen Formaenderung von Kristallen", Z. angew. Math. Mech., **8**, 161–185 (1928).
[8] W. Prager, "Discontinuous Solutions in the Theory of Plasticity", Courant Anniv. Vol., pp. 289–299, 1948.
[9] E. H. Lee, "On Stress Discontinuities in Plane Plastic Flow", Proc. 3rd Symposium on Appl. Math., McGraw-Hill Book Co., pp. 213–228, 1950.
[10] R. Hill, "The Mathematical Theory of Plasticity", Oxford Univ. Press, 1950.
[11] L. Prandtl, "Ueber die Eindringungsfestigkeit (Haerte) plastischer Baustoffe und die Festigkeit von Schneiden", Z. angew. Math. Mech., **1**, 15–20 (1921).
[12] A. Winzer and G. F. Carrier, "The Interaction of Discontinuity Surfaces in Plastic Fields of Stress", J. Appl. Mech., **15**, 261–264 (1948).

Brown University

(Received May 21, 1953)

Reprinted from JOURNAL OF MATHEMATICS AND PHYSICS
Vol. XXXIII, No. 2, July, 1954
*Printed in U.S.A.*

# LOADING PATHS AND THE INCREMENTAL STRAIN LAW*

G. H. HANDELMAN AND W. H. WARNER†

**1. Introduction.** This paper will be concerned with some properties of the stress-strain law for work hardening materials introduced by W. Prager [1]‡ and D. C. Drucker [2]. The law is of the type known as an *incremental strain* or *plastic flow* law by which it is meant that the differentials of strain are expressed as functions of the stresses, strains, and differentials of the stress. We shall also have occasion to refer to *total strain* or *plastic deformation* laws, in which the strains are given directly as functions of the stress.

The incremental strain laws considered here are very largely prescribed by a loading function $f(\sigma_{ij})$, where $\sigma_{ij}$ is the stress tensor. Loading is assumed to take place from a given state of stress under additional increments of stress if

$$(\partial f/\partial \sigma_{ij})d\sigma_{ij} > 0, \tag{1.1}$$

where the summation convention is implied. If the function $f$ depends only on the stress and not on the state of strain or strain history, the last condition may be written simply as $df > 0$. Only loading programs starting from an initially unstressed state will be considered.

The loading function is usually considered as a function of the nine variables $\sigma_{ij}$. It must be recognized that but six of these are independent. The nine components can be retained if the independent arguments of $f$ are considered to be $\frac{1}{2}(\sigma_{ij} + \sigma_{ji})$. In fact, it can be shown that the symmetry of the stress tensor restricts the loading function to be of this form. Furthermore, the assumption of no permanent volume change reduces the number of independent variables to five. These five may be considered to be the independent components of the stress deviator $s_{ij} = \sigma_{ij} - \frac{1}{3}\sigma_{kk}\delta_{ij}$. The latter statement is easily justified by a simple calculation.

It has been assumed tacitly that incremental strain laws of the Prager-Drucker type are in general non-integrable. In other words, they cannot be found as differentiated forms of suitably designed total strain laws. We shall prove that such is the case. Closely allied with this problem is a question originally raised by A. A. Ilyushin [3]—under what loading programs will a given incremental strain law agree with a given total strain law? Experimentally, this problem could be phrased as asking for a classification of tests under which a particular incremental strain law and a particular total strain law will give identical predictions for the relations between stress and strain. Furthermore, if two such laws agree under one type of testing program, will they agree under any other program?

* Work on this problem has been supported by the U. S. Air Force. The material in this paper was submitted as part of a thesis by the second author to the Carnegie Institute of Technology, June 1953, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

† Now at Brown University.

‡ Numbers in square brackets refer to references given at the end of the paper.

Ilyushin has given a partial, but not completely satisfactory, answer to the question. We will consider his results in the light of the general comments to be made here.

Finally, D. C. Drucker and G. N. White [4] have raised the problem of determining the conditions on the loading function for which the plastic work done will be independent of the loading path for isotropic work-hardening materials. They have given a particular example for which this is the case. We shall show that this problem is closely related to the others considered here and will discuss the most general form of loading function for which this property holds.

**2. Integration of the incremental strain laws.** Prager [1] and Drucker [2] have shown that, under rather broad assumptions, the most general incremental strain law must be of the following form. Corresponding to stress increments $d\sigma_{ij}$, there are strain increments $d\epsilon_{ij}$ which may be written as $d\epsilon_{ij} = d\epsilon'_{ij} + d\epsilon''_{ij}$. The elastic strain increments $d\epsilon'_{ij}$ are given by the differentiated Hooke's law:

$$d\epsilon'_{ij} = \frac{1+\nu}{E_0} d\sigma_{ij} - \frac{\nu}{E_0} d\sigma_{kk}\delta_{ij}, \tag{2.1}$$

where $\nu$ in Poisson's ratio and $E_0$ is Young's modulus. The non-elastic strain increment is given by

$$\begin{aligned} d\epsilon''_{ij} &= G \frac{\partial f}{\partial \sigma_{ij}} \frac{\partial f}{\partial \sigma_{kl}} d\sigma_{kl}, & \frac{\partial f}{\partial \sigma_{kl}} d\sigma_{kl} > 0 \\ &= 0, & \frac{\partial f}{\partial \sigma_{kl}} d\sigma_{kl} \leqq 0. \end{aligned} \tag{2.2}$$

Here $G$ is a scalar function which may depend upon stress, strain, and strain history.

We shall show first that the incremental strain law given by Eqs. (2.1) and (2.2) cannot be completely integrable; that is, this law does not represent the differentiated form of a reasonable total strain law. It will be assumed that the function $G$ and the loading function $f$ do not depend on the strains.

The condition of integrability can be expressed as [5]

$$\frac{\partial}{\partial \sigma_{kl}} (C_{ijmn}) = \frac{\partial}{\partial \sigma_{mn}} (C_{ijkl}) \tag{2.3}$$

for all $i$ and $j$. Here $C_{ijmn}$ is the coefficient of $d\sigma_{mn}$ in the relation

$$d\epsilon_{ij} = C_{ijmn} d\sigma_{mn}.$$

Since the elastic strain increment is integrable, Eq. (2.3) reduces to

$$\frac{\partial}{\partial \sigma_{kl}} \left( G \frac{\partial f}{\partial \sigma_{ij}} \frac{\partial f}{\partial \sigma_{mn}} \right) = \frac{\partial}{\partial \sigma_{mn}} \left( G \frac{\partial f}{\partial \sigma_{ij}} \frac{\partial f}{\partial \sigma_{kl}} \right)$$

for all $i, j, k, l, m, n$. Carrying out the indicated differentiations, and assuming the equality of mixed derivatives, we find (no summation)

$$\frac{\partial}{\partial \sigma_{kl}} \left( G \frac{\partial f}{\partial \sigma_{ij}} \right) \Big/ \frac{\partial}{\partial \sigma_{mn}} \left( G \frac{\partial f}{\partial \sigma_{ij}} \right) = \frac{\partial f}{\partial \sigma_{kl}} \Big/ \frac{\partial f}{\partial \sigma_{mn}}. \tag{2.4}$$

Equation (2.4) will be satisfied if [6]

$$G \frac{\partial f}{\partial \sigma_{ij}} = h_{ij}(f), \tag{2.5}$$

where the $h_{ij}$ are differentiable functions of $f$. Under the assumption of integrability, the incremental strain law becomes

$$d\epsilon''_{ij} = h_{ij}(f) \frac{\partial f}{\partial \sigma_{kl}} d\sigma_{kl} = h_{ij}(f)\, df,$$

which can be integrated in the form

$$\epsilon''_{ij} = H_{ij}(f). \tag{2.6}$$

We have assumed that the plastic strains vanish when $f = 0$, but this is not essential to the argument which follows.

We have thus shown that if the incremental strain law (2.2) is completely integrable, the corresponding plastic total strain law must be of the form (2.6). It is quite clear, however, that such a total strain law is completely unrealistic, for it states that the components of plastic strain will be the same for all states of stress for which the loading function $f$ has the same value. Consequently, if we consider a state of pure tension and a state of pure shear which lead to the same value of $f$, Eq. (2.6) requires that the corresponding plastic strains will be the same in both cases. Such conclusions must be rejected; and, therefore, the incremental strain law described by Eq. (2.2) cannot be completely integrable.

Since such an incremental strain law cannot agree completely with a total strain law for all possible loading programs, let us examine under what conditions such laws will agree. Such coincidence can best be studied in terms of those portions of stress space (a space in which the coordinates of a point are the components of the stress tensor) for which a given incremental law will agree with a given total strain law. Let us consider a total strain law of the form

$$\epsilon''_{ij} = \epsilon''_{ij}(\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{31}), \tag{2.7}$$

and assume that Eq. (3.5) can be solved for the stresses as functions of the plastic strains.* Thus we may equally well consider regions of coincidence in the plastic strainspace.

Let us assume that Eqs. (2.2) and (2.7) agree for certain states of stress and strain. Equation (2.2) can be written in the form

$$\frac{d\epsilon''_{11}}{\partial f/\partial \sigma_{11}} = \frac{d\epsilon''_{22}}{\partial f/\partial \sigma_{22}} = \cdots = \frac{d\epsilon''_{32}}{df/\partial \sigma_{32}} \; (= G\, df). \tag{2.8}$$

By means of the inverted forms of (2.7), the denominators in Eqs. (2.8) can be written as functions of $\epsilon''_{ij}$. Assuming $\sum_{i,j}(\partial f/\partial \sigma_{ij})^2 > 0$, we may integrate the system of ordinary differential equations (2.8) to obtain a curve in plastic strain space with $\epsilon''_{ij} = 0$ when $\sigma_{ij} = 0$ as our initial conditions. Using Eqs.

* The apparent discrepancy between five independent components of $\epsilon''_{ij}$ and six independent components of $\sigma_{ij}$ is illusory since the plastic strains are actually functions of the five components of the stress deviator.

(2.7), we obtain a corresponding curve in stress space. Unless all of the denominators appearing in the differential equations vanish (which may occur at the origin $\sigma_{ij} = 0$) there will be a unique curve passing through a given point of stress space. This means that emanating from the origin there is but one curve terminating at a given point in stress space along which Eqs. (2.2) and (2.7) agree. Alternatively, this means that for a predetermined state of stress there is at most one loading program starting with a state of zero stress and ending with the prescribed state of stress for which a given total strain law and a given incremental strain law will give identical predictions. The two laws must necessarily disagree for any other loading program. Hence in any experimental program which aims to study the predictions of two such laws it is important to determine the coincident loading programs and conduct tests which include paths other than these. Specific examples will be considered in the next section.

**3. Integration of some stress-strain laws for proportional loading.** As an application of the preceding discussion, let us consider the integration of the stress-strain laws given by Laning and Ilyushin [3] when the loading paths are those of proportional loading. The general law given by Hodge and Prager [7] may be treated in much the same way; the method used depends on the work of Handelman, Lin, and Prager [8].

The plastic part of the Laning flow law for work-hardening materials may be written as

$$\dot{e}''_{ij} = \frac{1}{4}\left(\frac{1}{G} - \frac{1}{G_0}\right)\frac{\dot{J}_2}{J_2}\, s_{ij}, \qquad J_2 > 0 \tag{3.1}$$

Here $e''_{ij}$ is the plastic part of the strain deviator, $s_{ij}$ is the stress deviator, $J_2$ is the second invariant of $s_{ij}$, $G$ is the tangent modulus in pure shear, and the dot represents differentiation with respect to some time-like parameter $\lambda$ on which the loading history depends. We assume that $e''_{ij} = 0$ when $s_{ij} = 0$.

Let the loading path be proportional:

$$s_{ij} = \bar{s}_{ij}\lambda, \tag{3.2}$$

where the $\bar{s}_{ij}$ are functions only of position. Assume that $G$ is independent of $\lambda$. We may then integrate Eq. (3.1). We have

$$\begin{aligned} J_2 &= \tfrac{1}{2}s_{ij}s_{ij} = \tfrac{1}{2}\bar{s}_{ij}\bar{s}_{ij}\lambda^2 = A\lambda^2 \\ \dot{J}_2 &= 2A\lambda. \end{aligned} \tag{3.3}$$

Hence

$$\dot{e}''_{ij} = \frac{1}{4}\left(\frac{1}{G} - \frac{1}{G_0}\right)\frac{2A\lambda}{A\lambda^2}\,\bar{s}_{ij}\lambda = \frac{1}{2}\left(\frac{1}{G} - \frac{1}{G_0}\right)\bar{s}_{ij}.$$

Integrating, we obtain

$$e''_{ij} = \frac{1}{2}\left(\frac{1}{G} - \frac{1}{G_0}\right)\bar{s}_{ij}\lambda = \frac{1}{2}\left(\frac{1}{G} - \frac{1}{G_0}\right)s_{ij}. \tag{3.4}$$

The resulting total strain law (3.4) is the plastic part of the Nadai deformation law.

*Added in proof.* The restriction that $G$ be independent of $\lambda$ is unnecessary, for if it were not we could still write

$\dot{e}''_{ij} = f(\lambda)\, \bar{s}_{ij}$ which could be integrated in the form

$e_{ij} = h(\lambda)\, s_{ij}$. The latter can in turn be written as

$e''_{ij} = g(J_2)\, s_{ij}$ which is the plastic part of the Nadai deformation law.

The incremental strain law of Ilyushin can be integrated to obtain his total strain law under the assumption of proportional loading. Again the dot indicates differentiation with respect to some time-like parameter $\lambda$. Here $J_2$ and $E_2$ are the second invariants of the stress and strain tensors, respectively.

The incremental law may be written as

$$(3.5) \qquad s_{ij} = \mu' \dot{e}_{ij}, \qquad \mu' = \sqrt{J_2/E_2}, \qquad J_2 = f\left(\int_0^\lambda \sqrt{\dot{E}_2}\, d\lambda, \sqrt{E_2}\right);$$

the total strain law, in the form

$$(3.6) \qquad s_{ij} = \mu'' e_{ij}, \qquad \mu'' = \sqrt{J_2/E_2}, \qquad J_2 = g(\sqrt{E_2}).$$

To have the two laws agree, we must have the two expressions for $J_2$ agree. If, in addition, we have proportional loading, we may integrate (3.5) to obtain (3.6). Let $\bar{s}_{ij} = \bar{s}_{ij} h(\lambda)$, $\bar{s}_{ij}$ independent of $\lambda$. To facilitate computation, let us take $h(\lambda)$ of the form

$$(3.7) \qquad h(\lambda) = \mu'' \exp\left\{\int_0^\lambda (\mu''/\mu')\, d\xi\right\}.$$

Then, from (3.5)

$$\dot{e}_{ij} = \frac{1}{\mu'} s_{ij} = \frac{\bar{s}_{ij}}{\mu'} h(\lambda);$$

and integration gives

$$e_{ij} = \bar{s}_{ij} \int_0^\lambda \left\{\frac{\mu''}{\mu'} \exp\left[\int_0^\xi \frac{\mu''}{\mu'}\, d\zeta\right]\right\} d\xi.$$

But

$$\frac{\mu''}{\mu'} \exp\left[\int_0^\xi \frac{\mu''}{\mu'}\, d\zeta\right] = \frac{d}{d\xi}\left\{\exp\left[\int_0^\xi \frac{\mu''}{\mu'}\, d\zeta\right]\right\};$$

hence

$$(3.8) \qquad \begin{aligned} e_{ij} &= \bar{s}_{ij} \exp\left[\int_0^\lambda \frac{\mu''}{\mu'}\, d\zeta\right] + C_{ij} \\ &= \bar{s}_{ij} \frac{h(\lambda)}{\mu''} + C_{ij} = \frac{s_{ij}}{\mu''}, \end{aligned}$$

where $C_{ij}$ has been set equal to zero under the assumption that $e_{ij} = 0$ when $s_{ij} = 0$. Thus (3.8) and (3.6) agree.

It can be shown that proportional loading is a necessary, as well as a sufficient, condition for the integration of (3.5) to obtain (3.6). Furthermore, Ilyushin gives as a criterion of integrability the condition that his "tensor of similitude" $\Pi = s_{ij}/\sqrt{3J_2}$ must be independent of $\lambda$: $\partial\Pi/\partial\lambda = 0$. A simple calculation shows that the latter condition is true if, and only if, the loading is proportional.

**4. Plastic work and the loading function.** White and Drucker [4] have raised the following question which is closely related to some of the problems considered in previous sections. If the increment of plastic work $dW''$ is defined for the incremental strain law (2.2) as

$$dW'' = \sigma_{ij} G \frac{\partial f}{\partial \sigma_{ij}} \frac{\partial f}{\partial \sigma_{kl}} d\sigma_{kl}, \tag{4.1}$$

under what conditions will $W''$ be independent of the loading path? White and Drucker have shown that if the loading function $f$ is homogeneous in the stresses and if $G$ is a function of $f$ alone, such will be the case. We shall discuss the most general form for the loading function if $W''$ is to be independent of the path. The latter condition is of importance, for instance, in the recent work of Y. Yamamoto [9], where basic minimum principles are developed, with path independence of the plastic work as an essential feature of their proof. Hill [10] has also considered related questions.

We note that this requirement is equivalent to (4.1) being completely integrable. This implies that [5]

$$\frac{\partial}{\partial \sigma_{pq}}\left(G\sigma_{ij}\frac{\partial f}{\partial \sigma_{ij}}\frac{\partial f}{\partial \sigma_{kl}}\right) = \frac{\partial}{\partial \sigma_{kl}}\left(G\sigma_{ij}\frac{\partial f}{\partial \sigma_{ij}}\frac{\partial f}{\partial \sigma_{pq}}\right),$$

which, after some simplification based on the continuity of the derivatives of $f$, becomes

$$\frac{\partial}{\partial \sigma_{kl}}\left(G\sigma_{ij}\frac{\partial f}{\partial \sigma_{ij}}\right) \Big/ \frac{\partial}{\partial \sigma_{pq}}\left(G\sigma_{mn}\frac{\partial f}{\partial \sigma_{mn}}\right) = \frac{\partial f}{\partial \sigma_{kl}} \Big/ \frac{\partial f}{\partial \sigma_{pq}}. \tag{4.2}$$

Equation (4.2) has the solution [6]

$$G\sigma_{ij}\frac{\partial f}{\partial \sigma_{ij}} = g(f), \tag{4.3}$$

where $g(f)$ is an arbitrary differentiable function of $f$. If we assume, further, that $G$ is a function of $f$ only, then Eq. (4.3) can be replaced by

$$\sigma_{ij}\frac{\partial f}{\partial \sigma_{ij}} = h(f), \tag{4.4}$$

where $h(f)$ is an arbitrary differentiable function of $f$. The condition that $f$ be a homogeneous function in the stresses of degree $n$, as given by White and Drucker, is a particular solution of Eq. (4.4), for by Euler's Theorem we have in this case

$$\sigma_{ij}\frac{\partial f}{\partial \sigma_{ij}} = nf.$$

The most general solution of Eq. (4.4) may be studied by the usual method of solving a quasi-linear non-homogeneous first order partial differential equation. Let

$$V = V(\sigma_{ij}, f) = 0$$

be some relation defining $f$ implicitly as a function of the stresses. Then

$$\frac{\partial V}{\partial \sigma_{ij}} + \frac{\partial V}{\partial f}\frac{\partial f}{\partial \sigma_{ij}} = 0;$$

or, if it is assumed that $\partial V/\partial f \neq 0$,

$$\frac{\partial f}{\partial \sigma_{ij}} = -\frac{\partial V}{\partial \sigma_{ij}} \Big/ \frac{\partial V}{\partial f}.$$

Substituting these relations in Eq. (4.4), we find

$$\sigma_{ij}\frac{\partial V}{\partial \sigma_{ij}} + h(f)\frac{\partial V}{\partial f} = 0. \tag{4.5}$$

To this we adjoin the following system of ordinary differential equations

$$d\sigma_{11}/\sigma_{11} = \cdots = d\sigma_{22}/\sigma_{22} = df/h(f),$$

the nine independent integral curves of which are ($\sigma_{11} \neq 0$)

$$\exp\left[\int^f d\xi/h(\xi)\right] = k\sigma_{11}, \qquad k = \text{const.}$$

$$\sigma_{ij} = c_{ij}\sigma_{11}, \qquad c_{ij} \text{ constants, } i \text{ and } j \text{ not both } 1.$$

Hence the general solution of Eq. (4.5) is of the form

$$V = V\left(\sigma_{ij}/\sigma_{11}, \quad \frac{1}{\sigma_{11}}\exp\left[\int^f d\xi/h(\xi)\right]\right).$$

where $V$ is an arbitrary function possessing the necessary differentiability properties. Setting $V = 0$ and solving for the last variable, we have

$$\frac{1}{\sigma_{11}}\exp\left[\int^f d\xi/h(\xi)\right] = \Phi(\sigma_{ij}/\sigma_{11}),$$

or

$$H(f) = \exp\left[\int^f d\xi/h(\xi)\right] = \sigma_{11}\Phi\left(\frac{\sigma_{22}}{\sigma_{11}}, \cdots, \frac{\sigma_{22}}{\sigma_{11}}\right)$$

where $\Phi$ is an arbitrary function of the variables indicated. Hence $H(f)$ is homogeneous of degree one in the stresses. If $H$ has an inverse, we see that

$$f(\sigma_{ij}) = H^{-1}\left(\sigma_{11}\Phi\left[\frac{\sigma_{22}}{\sigma_{11}}, \cdots, \frac{\sigma_{22}}{\sigma_{11}}\right]\right). \tag{4.6}$$

Equation (4.6) then states that the plastic work will be independent of the path if the loading function $f$ is any function of a function homogeneous of degree one in the stresses, provided that the former function has an inverse and is suffi-

ciently differentiable. We see that the White and Drucker solution is a special case of Eq. (4.6) since any homogeneous function $g$ of degree $n$ may be written as

$$g = \sigma_{11}^n k\left(\frac{\sigma_{22}}{\sigma_{11}}, \cdots, \frac{\sigma_{22}}{\sigma_{11}}\right).$$

The present solution allows loading functions which are themselves non-homogeneous in the stresses; for instance, if $\phi$ is homogeneous of degree one in the stresses, then

$$f = \phi + \phi^2$$

is not homogeneous but admits of a path-independent plastic work function.

**5. Conclusion.** At least two further questions connected with this study remain to be answered. We have shown that a given incremental strain law and a given total strain law will agree exactly only for certain loading programs. In fact, many of the well known laws agree only for proportional loading. In many practical problems of stress analysis, it is a convenient mathematical simplification to use a total strain law rather than an incremental law. The question that remains then is to see if it is possible to obtain estimates of the error committed when a total strain law is used in place of an incremental law for a loading history during which the two laws do not coincide.

Secondly, we have given a general criterion on the loading function, Eq. (4.6), which insures that the plastic work will be independent of the loading path. It remains to be seen whether loading functions can be constructed which will be otherwise physically reasonable and still conform to this criterion.

## REFERENCES

1. W. Prager, *Recent development in the mathematical theory of plasticity*, J. App. Phys., **20**, 235–241 (1949).
2. D. C. Drucker, *Some implications of work hardening and ideal plasticity*, Quart. Appl. Math., **7**, 411–418 (1950).
3. A. A. Ilyushin, *Relation between the theory of St. Venant-Levy-Mises and the theory of small elastic-plastic deformations*, Prik. Mat. i Mekh., **9**, 207–218 (1945); Brown University Translation (1945).
4. G. N. White and D. C. Drucker, *Effective stress and effective strain in relation to stress theories of plasticity*, J. App. Phys., **21**, 1013–1021 (1950).
5. Goursat-Hedrick, *Mathematical Analysis*, v. 2, part 2, Ginn and Co., New York (1917), p. 51.
6. Goursat-Hedrick, *Ibid.*, v. 1, p. 57.
7. P. Hodge and W. Prager, *A variational principle for plastic materials with strain-hardening*, J. Math. and Phys., **27**, 1–10 (1948).
8. G. H. Handelman, C. C. Lin, and W. Prager, *On the mechanical behavior of metals in the strain-hardening range*, Quart. Appl. Math., **4**, 397–407 (1947).
9. Y. Yamamoto, *Variational principles of equilibrium of an elasto-plastic body*, Quart. Appl. Math., **10**, p. 215 (1952).
10. R. Hill, *The mathematical theory of plasticity*, Oxford, 1950, pp. 37–38.

Carnegie Institute of Technology

(Received September 28, 1953)

# THE OPERATIONAL CALCULUS OF LEGENDRE TRANSFORMS*

By R. V. Churchill

**1. Introduction.** The sequence of numbers $f(n)$ defined by the equation

$$f(n) = \int_{-1}^{1} F(x)P_n(x)\,dx \qquad n = 0, 1, 2, \cdots, \tag{1}$$

where $P_n(x)$ denotes the Legendre polynomial of degree $n$, is the Legendre transform of the function $F(x)$. The integral transformation here will be represented by the symbol $T\{F(x)\}$. For functions $F(x)$ satisfying well-known conditions on the interval $-1 \leqq x \leqq 1$ the inverse of this transformation is represented by the expansion of $F(x)$ in series of the Legendre polynomials,

$$F(x) = \sum_{n=0}^{\infty} (n + \tfrac{1}{2})f(n)P_n(x) = T^{-1}\{f(n)\} \qquad (-1 < x < 1). \tag{2}$$

Let $R[F]$ denote the differential form

$$R[F(x)] = \frac{d}{dx}\left[(1 - x^2)\frac{d}{dx}F(x)\right]. \tag{3}$$

When the integral $T\{R[F]\}$ is integrated successively by parts and $-n(n + 1)P_n(x)$ is substituted for $R[P_n(x)]$ in accordance with Legendre's differential equation, the following result is easily obtained.

Theorem 1: Let $F(x)$ denote a function that satisfies these conditions: $F'(x)$ is continuous and $F''(x)$ is bounded and integrable over each interval interior to the interval $-1 < x < 1$; $T\{F(x)\}$ exists and

$$\lim_{x\to\pm 1} (1 - x^2)F(x) = \lim_{x\to\pm 1} (1 - x^2)F'(x) = 0.$$

Then $T\{R[F(x)]\}$ exists and

$$T\{R[F(x)]\} = -n(n + 1)f(n) \qquad (n = 0, 1, 2, \cdots). \tag{4}$$

Formula (4) represents the basic operational property of the Legendre transformation $T$ under which the *differential* operation $R[F]$ defined by equation (3) is replaced by the *algebraic* operation $-n(n + 1)f(n)$.

The established forms of operational calculus for solving problems in differential equations are based on integral transformations of Fourier type, transformations whose kernels are exponential functions or linear combinations of such functions. These transformations consist of the various Laplace and Fourier integral transformations. Transforms of the other types, including Legendre transforms [1], have been recognized as bases for other forms of operational calculus [2, 3, 4], but to date the operational methods have not been developed beyond the stage of applying the basic operational property corresponding to formula (4) and the application of the inverse transformation. The type of

* Sponsored by the Office of Ordnance Research, U. S. Army, under Contract DA-20-018-ORD-12916.

boundary value problem that can be reduced by an integral transformation is of course governed by the kernel and the interval of integration [5].

A convolution property, one that gives directly the image of the operation of taking the product of the transforms of two functions, is now known for Legendre transforms [6]. Additional operational properties will be noted here and a short table of Legendre transforms will be developed. The operational calculus will then be illustrated by applying it to some classical boundary value problems in partial differential equations.

The application to the third boundary value problem for the potential inside a sphere leads to a simple expression for the solution of that problem in terms of the solution of a corresponding Dirichlet problem. This formula, together with the accompanying extension of the Poisson integral formula to the problem of the third kind, should be known; but a search of the literature and inquiry among colleagues has not yet revealed these formulas.

**2. Operational properties.** If each of the functions $R[F(x)]$ and $F(x)$ satisfy the sufficient conditions stated for the validity of formula (4) then the transform of the iterated differential form $R[R[F]]$ can be written at once as

$$T\{R^2[F(x)]\} = n^2(n+1)^2 f(n). \tag{5}$$

Similarly for iterations of higher order.

The substitution of $(n+\frac{1}{2})^2 - \frac{1}{4}$ for $n(n+1)$ in formula (4) leads to this modification of that operational property:

$$(n+\tfrac{1}{2})^2 f(n) = T\{\tfrac{1}{4}F(x) - R[F(x)]\}. \tag{6}$$

It should be noted that under the substitution $x = \cos\theta (0 \leqq \theta \leqq \pi)$ our transformation

$$T\{F(\cos\theta)\} = \int_0^\pi F(\cos\theta) P_n(\cos\theta) \sin\theta \, d\theta = f(n)$$

ransforms the differential form

$$R[F(\cos\theta)] = \frac{1}{\sin\theta}\frac{d}{d\theta}\left[\sin\theta \frac{d}{d\theta} F(\cos\theta)\right] \tag{7}$$

nto $-n(n+1)f(n)$, according to formula (4).

Consider now the transform of the function $R^{-1}[F]$, where $R^{-1}$ is the inverse iof the differential operator $R$. Let $Y(x)$ denote the function $R^{-1}[F(x)]$; then t$Y(x)$ is a solution of the differential equation

$$R[Y(x)] = F(x). \tag{8}$$

Suppose that $F(x)$ is a function of bounded variation on each interval $|x| \leqq |x_1| < 1$, and that

$$\int_{-1}^{1} F(x)\, dx = 0; \tag{9}$$

that is, $f(0) = 0$. The first integral of equation (8),

$$(1 - x^2)Y'(x) = \int_{-1}^{x} F(t)\, dt$$

is then a continuous function of $x(|x| \leqq 1)$ with limit zero as $x \to \pm 1$. The second integral

$$Y(x) = \int_0^x \frac{1}{1 - s^2} \int_{-1}^{s} F(t)dt\, ds + C = R^{-1}[F], \tag{10}$$

where $C$ is an arbitrary constant, is continuous when $|x| < 1$ and it is easy to show that $(1 - x^2)Y(x)$ vanishes as $x \to \pm 1$; in fact $|Y(x)|$ is of the order of $(1 - x^2)^{-k}$ for each positive constant $k < 1$ as $x \to \pm 1$ and hence $T\{Y\}$ exists. According to Theorem 1 and equation (8) then

$$T\{R[Y]\} = -n(n + 1)T\{Y\} = f(n);$$

thus

$$T\{R^{-1}[F]\} = -\frac{f(n)}{n(n + 1)} \qquad (n = 1, 2, \cdots). \tag{11}$$

The value of the transform of $R^{-1}[F]$ at $n = 0$ depends on the value assigned to the constant $C$; if $F(x)$ is an odd function the value is easily shown to be $2C$. The operational property concerning $R^{-1}$ can be stated as follows.

THEOREM 2: Let $F(x)$ denote a function of bounded variation in each subinterval of the interval $-1 < x < 1$ and let $f(0) = 0$. Then $f(n)$ exists and for each constant C,

$$T^{-1}\left\{\frac{-f(n)}{n(n + 1)}\right\} = R^{-1}[F(x)] = \int_0^x \frac{1}{1 - s^2} \int_{-1}^{s} F(t)dt\, ds + C \tag{12}$$

$$(n = 1, 2, \cdots).$$

The convolution property can be stated as follows [6].

THEOREM 3: Let $F(x)$ and $G(x)$ denote bounded integrable functions on the interval $-1 \leqq x \leqq 1$. Then the product $f(n)g(n)$ of their Legendre transforms is the transform of the function $H(x)$; that is,

$$T^{-1}\{f(n)g(n)\} = H(x) \tag{13}$$

where $H(x)$ is described by any one of these formulas:

$$H(\cos\theta) = 1/\pi \int_0^{\pi} F(\cos\theta') \sin\theta' \int_0^{\pi} G(\cos\lambda)\, d\phi\, d\theta' \tag{14}$$

where

$$\cos\lambda = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\phi;$$

$$H(\cos\theta) = 1/\pi \int_0^{\pi/2} \sin\phi \int_0^{2\pi} F[\sin\phi \sin(\beta - \tfrac{1}{2}\theta)]G[\sin\phi \sin(\beta + \tfrac{1}{2}\theta)]\, d\beta\, d\phi; \tag{15}$$

$$H(x) = 1/\pi \iint_{E_x} F(y)G(z)(1 - x^2 - y^2 - z^2 + 2xyz)^{-1/2}\, dy\, dz \tag{16}$$

where, for each fixed $x(-1 < x < 1)$, $E_x$ is the region interior to the ellipse

$$y^2 + z^2 - 2xyz = 1 - x^2.$$

Formula (14) has a geometrical interpretation. Let $(r, \phi, \theta)$ denote spherical polar coordinates. If $P$ denotes any point $(1, 0, \theta)$ on the boundary of a hemispherical surface of unit radius and $P'$ any point $(1, \phi, \theta')$ on that surface, then according to the cosine law for spherical triangles $\lambda$ is the arc $PP'$ of a great circle. If $S$ denotes the hemisphere $r = 1, 0 \leqq \theta' \leqq \pi, 0 \leqq \phi \leqq \pi$ then formula (14) can be written

$$H(\cos\theta) = 1/\pi \iint_S F(\cos\theta')G(\cos\lambda)\, dS; \tag{17}$$

thus, $H(\cos\theta)$ is twice the mean value of $F(\cos\theta')G(\cos\lambda)$ over the hemisphere.

For $\theta = 0$ and $\theta = \pi$ formula (14) reduces to

$$H(1) = \int_{-1}^{1} F(t)G(t)\, dt, \qquad H(-1) = \int_{-1}^{1} F(t)G(-t)\, dt, \tag{18}$$

respectively. In view of formula (2) and the fact that $P_n(1) = 1$ the first of equations (18) is seen to be the Parseval relation for the orthogonal set of functions $P_n(x)$.

THEOREM 4: If $T\{F(x)\}$ and $T\{G(x)\}$ exist then

$$T\{C_1F(x) + C_2G(x)\} = C_1f(n) + C_2g(n), \tag{19}$$

where $C_1$ and $C_2$ are constants; also

$$T\{F(-x)\} = (-1)^n f(n). \tag{20}$$

These are obvious properties of the transformation since $T$ is clearly linear and since $P_n(-x) = (-1)^n P_n(x)$. When $G(x) = 1$ then $g(n) = 0$ $(n = 1, 2, \cdots)$ and $g(0) = 2$; according to equation (19) then, if $C$ is a constant,

$$T\{F(x) + C\} = f(n) \qquad \text{when } n = 1, 2, \cdots,$$
$$= f(0) + 2C \qquad \text{when } n = 0.$$

The images under $T$ of the operations of differentiation and indefinite integration involve finite differences. Let $F(x)$ be a sectionally continuous function on the interval $-1 \leqq x \leqq 1$ and let $G(x)$ denote the continuous function

$$G(x) = \int_{-1}^{x} F(t)\, dt.$$

Then

$$f(n) = \int_{-1}^{1} G'(x)\, P_n(x)\, dx = G(1) - \int_{-1}^{1} G(x)P_n'(x)\, dx$$

and it follows from a differential recurrence relation for $P_n(x)$ that

$$f(n-1) - f(n+1) = \int_{-1}^{1} G(x)[P'_{n+1}(x) - P'_{n-1}(x)]\,dx$$

$$= (2n+1)g(n) \qquad (n = 1, 2, \cdots); \tag{21}$$

also, $f(0) = G(1)$ and since $P_1(x) = x, f(1) = G(1) - g(0)$ so that

$$f(0) - f(1) = g(0).$$

The solution of the difference equation (21) for $g(n)$ and $f(n)$ in turn leads to these conclusions:

THEOREM 5: If $F(x)$ is sectionally continous then

$$T\left\{\int_{-1}^{x} F(t)\,dt\right\} = \frac{f(n-1) - f(n+1)}{2n+1} \tag{22}$$

where, for $n = 0, f(n-1)$ is to be replaced by $f(0)$. If $G(x)$ is continuous and $G'(x)$ sectionally continuous and if $G(-1) = 0$ then

$$T\{G'(x)\} = G(1) - (2n-1)g(n-1) - (2n-5)g(n-3) - \cdots - g(0)$$

$$(n = 1, 3, \cdots),$$

$$= G(1) - (2n-1)g(n-1) - (2n-5)g(n-3) \tag{23}$$

$$- \cdots - 3g(1) \qquad (n = 2, 4, \cdots),$$

$$= G(1) \qquad (n = 0).$$

Other recurrence formulas for $P_n(x)$ lead to operational properties of the transforms, but none seem simple or promising.

**3. Transforms of particular functions.** A short table of Legendre transforms is presented here. Some methods of computing transforms in the table will now be indicated.

It follows at once from the orthogonality properties of the Legendre polynomials that when $F(x) = P_m(x)(m = 0, 1, 2, \cdots)$ then

$$f(n) = 0(n \neq m), f(m) = 1/(m + \tfrac{1}{2}).$$

From the well-known representation of the function $G(x) = x^m$ in a finite series of Legendre polynomials of degree $m$ and lower it follows that $g(n) = 0$ when $n > m$. The formula for $g(n)$ when $n \leqq m$ is not simple (cf. [7]). If $F(x)$ is any polynomial of degree $m$ then $f(n) = 0$ when $n > m$.

The function $F(x) = \log(1 - x)$ satisfies the condition $R[F] = -1$, where $R$ is the differential operator involved in Theorem 1. But $F'(x)$ does not satisfy the limit condition at $x = 1$ in that theorem. Integration by parts, however, gives the equation

$$T\{R[F]\} = -(1+x)P_n(x)\,|_{-1}^{1} - (1-x^2)\log(1-x)P'_n(x)\,|_{-1}^{1} - n(n+1)f(n);$$

that is,

$$T\{-1\} = -2 - n(n+1)f(n) = 0 \qquad (n = 1, 2, \cdots).$$

TABLE OF LEGENDRE TRANSFORMS

| | $f(n)$ | $(n = 0, 1, 2, \cdots)$ | $F(x)$ $(-1 < x < 1)$ |
|---|---|---|---|
| 1. | $\frac{1}{m + 1/2}$ if $n = m$; | 0 if $n \neq m$ | $P_m(x)$ $(m = 0, 1, 2, \cdots)$ |
| 2. | $\frac{1}{n(n+1)}$ if $n > 0$; | $1 - \log 2$ if $n = 0$ | $-\frac{1}{2}\log(1 - x)$ |
| 3. | $\frac{(-1)^n}{n(n+1)}$ if $n > 0$; | $1 - \log 2$ if $n = 0$ | $-\frac{1}{2}\log(1 + x)$ |
| 4. | $a^n$ | $(\lvert a \rvert < 1)$ | $\frac{1}{2}(1 - a^2)(1 - 2ax + a^2)^{-3/2}$ |
| 5. | $\frac{a^n}{n}$ if $n > 0$; | 2 if $n = 0$ $(\lvert a \rvert \leqq 1)$ | $(1 - 2ax + a^2)^{-1/2} - \frac{1}{2}\log\frac{1 - ax + (1 - 2ax + a^2)^{1/2}}{2}$ |
| 6. | $\frac{a^n}{n + 1/2}$ | $(\lvert a \rvert \leqq 1)$ | $(1 - 2ax + a^2)^{-1/2}$ |
| 7. | $\frac{a^{n+1}}{n+1}$ | $(\lvert a \rvert \leqq 1)$ | $a(1 - 2ax + a^2)^{-1/2} - \frac{1}{2}\log\frac{a - x + (1 - 2ax + a^2)^{1/2}}{1 - x}$ |
| 8. | $\frac{a^{n+1}}{(n + 1/2)(n+1)}$ | $(\lvert a \rvert \leqq 1)$ | $\log\frac{a - x + (1 - 2ax + a^2)^{1/2}}{1 - x}$ |
| 9. | $\frac{a^n}{n(n + 1/2)}$ if $n > 0$; | 0 if $n = 0$ | $-\log\frac{1 - ax + (1 - 2ax + a^2)^{1/2}}{2}$ $(\lvert a \rvert \leqq 1)$ |
| 10. | $\frac{a^{n+c}}{(n + 1/2)(n + c)}$ | $(c > 0, \lvert a \rvert \leqq 1)$ | $\int_0^a \frac{t^{c-1}\,dt}{(1 - 2tx + t^2)^{1/2}}$ |
| 11. | $\frac{P_{n-1}(0) - P_{n+1}(0)}{2n + 1}$ if $n > 0$; | 1 if $n = 0$ | 0 if $x < 0$; 1 if $x > 0$ |

Direct integration gives $f(0) = 2\log 2 - 2$; hence

$$T\{\log(1-x)\} = -2/n(n+1) \qquad (n > 0),$$
$$= 2\log 2 - 2 \qquad (n = 0).$$

The transform of $\log(1+x)$ follows from Theorem 4.

From the uniformly convergent expansion of the generating function

$$(1 - 2tx + t^2)^{-\frac{1}{2}} = \sum_{n=0}^{\infty} t^n P_n(x) \qquad (|t| < 1)$$

it follows that

$$\int_{-1}^{1} (1 - 2tx + t^2)^{-\frac{1}{2}} P_n(x)\, dx = t^n/(n + \tfrac{1}{2}) \qquad (|t| < 1). \tag{24}$$

When $t > 0$ the integrand here is a continuous function of $t$ and $x$ except at $x = t = 1$ and it is dominated by the function $(1 - x^2)^{-\frac{1}{2}}$. Hence the integral is uniformly convergent with respect to $t$ up to $t = 1$ and it represents a continuous function of $t$ there. Thus the transformation is valid when $t = 1$; that is,

$$T\{(1-x)^{-\frac{1}{2}}\} = \sqrt{2}/(n + \tfrac{1}{2}).$$

The transform of $(1+x)^{-\frac{1}{2}}$ follows from Theorem 4.

Differentiation of both members of equation (24) with respect to the parameter $t$ followed by multiplication by $t$ leads to the equation

$$\tfrac{1}{2}T\{(1 - 2tx + t^2)^{-3/2}(2tx - 2t^2)\} = nt^n/(n + \tfrac{1}{2}) \qquad (|t| < 1).$$

Since $2tx - 2t^2 = -(1 - 2tx + t^2) + 1 - t^2$ it follows with the aid of formula (24) that

$$T\{(1 - 2tx + t^2)^{-3/2}\} = 2t^n/(1 - t^2) \qquad (|t| < 1). \tag{25}$$

A continuation of the process leads to a slightly involved expression for $T^{-1}\{nt^n\}$ when $|t| < 1$.

The transformation

$$T\left\{\log \frac{a - x + (1 - 2ax + a^2)^{\frac{1}{2}}}{1 - x}\right\} = \frac{a^{n+1}}{(n + \frac{1}{2})(n+1)} \qquad (|a| \leqq 1)$$

can be verified by integrating both members of equation (24) with respect to $t$. From the partial fractions expansion of the coefficient of $a^{n+1}$ here and from formula (24) with $t = a (|a| \leqq 1)$ it follows that

$$T\left\{\frac{2a}{(1 - 2ax + a^2)^{\frac{1}{2}}} - \log \frac{a - x + (1 - 2ax + a^2)^{\frac{1}{2}}}{1 - x}\right\} = \frac{2a^{n+1}}{n+1} \qquad (|a| \leqq 1).$$

From equation (24) it follows that

$$\int_{-1}^{1} [t^{-1}(1 - 2tx + t^2)^{-\frac{1}{2}} - t^{-1}] P_n(x)\, dx = t^{n-1}/(n + \tfrac{1}{2}) \qquad (n = 1, 2, \cdots)$$
$$= 0 \qquad (n = 0),$$

and the integration of this equation with respect to $t$ leads to the transformation

$$T\left\{-\log\frac{1-ax+(1-2ax+a^2)^{\frac{1}{2}}}{2}\right\}=\frac{a^n}{n(n+\frac{1}{2})} \qquad (n=1,2,\cdots;\,|a|\leqq 1)$$

$$=0 \qquad (n=0;\,|a|\leqq 1).$$

From the representation of the coefficient of $a^n$ here in partial fractions the inverse transform of $a^n/n$ follows with the aid of formula (24).

Transform No. 10 in the table can be found by multiplying equation (24) by $t^{-1}$ and integrating with respect to $t$. Transforms of step functions, illustrated by No. 11 in the table, are easily written with the aid of a well-known integration formula for $P_n(x)$. Combinations and special cases of the transforms in the table, as well as applications of the operational properties of the preceding sections, lead to a considerable extension of the list of transforms.

**4. Applications. Dirichlet problem for sphere.** Let $V(r, \cos\theta)$ denote the potential function interior to the unit sphere when the potential on the surface $r = 1$ is a prescribed function $F(\cos\theta)$ of the spherical coordinate $\theta$ only $(0 \leqq \theta \leqq \pi)$. The interior is free from sources so that $V$ satisfies Laplace's equation

$$\nabla^2 V=\frac{1}{r}(rV)_{rr}+\frac{1}{r^2\sin\theta}(\sin\theta V_\theta)_\theta=0$$

when $r < 1$. If $x = \cos\theta$ this equation can be written

$$r(rV)_{rr}+[(1-x^2)V_x]_x=0 \qquad (r<1). \tag{26}$$

The function $V$ is also bounded in the region and satisfies the boundary condition

$$V(1,x)=F(x) \qquad (-1\leqq x\leqq 1). \tag{27}$$

To solve this boundary value problem formally by means of the operational calculus developed in the foregoing sections the problem is written in terms of the transforms

$$v(r,n)=T\{V(r,x)\},\ f(n)=T\{F(x)\} \qquad (n=0,1,2,\cdots)$$

with respect to the variable $x$. Let the operator $T$ be applied to the members of equation (26) and (27) and let the order of differentiating $V$ with respect to $r$ and then integrating with respect to $x$ be reversed. In view of the basic operational property (Theorem 1) it follows that $v(r, n)$ satisfies the conditions

$$r\,d^2(rv)/dr^2-n(n+1)v=0, \qquad v(1,n)=f(n),$$

and $v(r, n)$ is bounded when $0 \leqq r \leqq 1$.

The solution of the simple problem in $v$ here is

$$v(r,n)=f(n)r^n$$

and the inversion formula (2) can be used to represent $V(r, x)$ by an infinite series in the functions $r^nP_n(x)$, the form of the solution that would be obtained by separating variables in the boundary value problem (26), (27). The function

$V(r, x)$ can be written in closed form, however, with the aid of the convolution property (Theorem 3) and the inverse transform of $r^n$ (No. 4 of the table). According to formula (14) for the inverse transform of the product of two transforms the inverse transform of $v(r, n)$ is

$$V(r, \cos\theta) = \frac{1 - r^2}{2\pi} \int_0^\pi F(\cos\theta') \sin\theta' \int_0^\pi (1 + r^2 - 2r\cos\lambda)^{-3/2}\, d\phi\, d\theta' \qquad (r < 1), \tag{28}$$

where

$$\cos\lambda = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\phi.$$

This is the well-known Poisson integral formula for the potential inside a sphere in this special case in which the potential is a function of $r$ and $\theta$ only.

When $F(x)$ is a linear combination of any of the functions listed among the first nine items of the table, then $f(n)r^n$ can be written as a linear combination of the transforms listed in the table and the function $V$ can be written in a simple form free from integrals.

In the corresponding problem for the potential $W(r, \cos\theta)$ in the region exterior to the sphere $r = 1$, the solution of the transformed problem is

$$w(r, n) = (1/r)f(n)(1/r^n) \qquad (r \geqq 1).$$

Hence

$$W(r, \cos\theta) = (1/r)V(1/r, \cos\theta) \tag{29}$$

where $V$ denotes the above potential function for the interior region. The Poisson integral formula for the potential $W$ follows at once from formulas (29) and (28).

When the potential is not independent of the spherical coordinate $\phi$ the Laplacian involves a differential operator with respect to $\cos\theta (x = \cos\theta)$ that is more involved than the operator $R$ of equation (3) and the transformation $T$ does not eliminate derivatives with respect to $x$ from the Laplacian.

**5. Neumann problem for sphere.** Let $U(r, \cos\theta)$ denote the potential in the region interior to the sphere $r = 1$ when the normal derivative is a prescribed function $F(\cos\theta)$, so that

$$\nabla^2 U = 0 \ (r < 1), \qquad U_r(1, x) = F(x), \tag{30}$$

where $x = \cos\theta$, and $U$ is bounded when $0 \leqq r \leqq 1$. Then $u(r, n)$, the transform of $U(r, x)$, is bounded and formally satisfies the conditions

$$r(ru)'' - n(n + 1)u = 0, \qquad u'(1, n) = f(n) \qquad (n = 0, 1, 2, \cdots)$$

where the primes denote differentiation with respect to $r$.

When $n = 0$ the bounded solution of the differential equation here is $u(r, 0) = C$, where $C$ is a constant. Hence $f(0) = 0$ if the boundary condition is to be satisfied; thus

$$\int_0^\pi F(\cos\theta) \sin\theta\, d\theta = 0;$$

that is, the mean value of $F(\cos\theta)$ over the surface is zero, a well-known necessary condition for the solution of the Neumann problem.

The transformed problem has the solution

$$u(r, n) = f(n)r^n/n \qquad (n = 1, 2, \cdots) \tag{31}$$
$$= C \qquad (n = 0)$$

where $C$ is an arbitrary constant. Thus the potential $U(r, x)$ is determined up to an additive constant. It can be expressed in a series of the functions $r^nP_n(x)$ with the aid of the inversion formula (2); but it can be written in the closed form

$$U(r, \cos\theta) = \frac{1}{\pi}\int_0^\pi F(\cos\theta')\sin\theta' \int_0^\pi \{(1 + r^2 - 2r\cos\lambda)^{-\frac{1}{2}} \tag{32}$$
$$- \frac{1}{2}\log[1 - r\cos\lambda + (1 + r^2 - 2r\cos\lambda)^{\frac{1}{2}}]\}\, d\phi'\, d\theta' + C,$$

where

$$\cos\lambda = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\phi',$$

with the aid of the convolution formula (14) and transform No. 5 in the table.

Let $Q$ denote any point $(1, \phi', \theta')$ on the hemisphere $S$, $r = 1$, $0 \leqq \phi' \leqq \pi$, $0 \leqq \theta' \leqq \pi$, and let $P$ denote any point $(r, 0, \theta)$ of the region bounded by the boundary circle of $S$. Then the above variable $\lambda$ represents the angle between the radii through $P$ and $Q$ and the length of the line segment $PQ$ is

$$\rho = (1 + r^2 - 2r\cos\lambda)^{\frac{1}{2}}.$$

With the aid of formula (17) for the convolution, an alternate form of equation (32) now can be written as

$$U(r, \cos\theta) = \frac{1}{2\pi}\iint_S F(\cos\theta')[(2/\rho) - \log(1 + \rho - r\cos\lambda)]\, dS + C. \tag{33}$$

This is a known form [8] of the Poisson integral solution of the Neumann problem for the sphere.

From equation (31) it follows that

$$ru'(r, n) = f(n)r^n = v(r, n) \qquad (n = 1, 2, \cdots),$$

where $v(r, n)$ is the transform of the Dirichlet problem (26), (27), with the same function $F(x)$ there as here. Thus

$$u(r, n) = \int_0^r v(r', n)\, dr'/r' \qquad (n = 1, 2, \cdots),$$
$$= C \qquad (n = 0)$$

and the inverse transformation gives the expression

$$U(r, \cos\theta) = \int_0^r V(r', \cos\theta)dr'/r' + C \tag{34}$$
$$= \int_0^1 V(rt, \cos\theta)\, dt/t + C$$

for the solution of the Neumann problem (30) in terms of the solution of the Dirichlet problem (26), (27). This relationship between the solutions can be verified directly and it *is not necessary to assume that the potential functions and the function $F$ are independent of the coordinate $\phi$.*

**6. Problem of third type for sphere.** Now let $U(r, \cos\theta)$ represent the bounded potential function for the region interior to the sphere $r = 1$ satisfying the mixed type of boundary condition at the surface. Specifically,

$$\nabla^2 U = 0 \ (r < 1), \qquad U_r(1, x) + (k+1)U(1, x) = F(x) \tag{35}$$

where $x = \cos\theta$ and $k + 1$ is a positive constant. The function $U$ can be interpreted as the steady-state temperatures at points in a solid sphere which is subjected to linear heat transfer at its surface into surroundings whose temperature is proportional to $F(\cos\theta)$.

The transform $u(r, n)$ of $U(r, x)$ is the bounded function that satisfies the conditions

$$r(ru)'' - n(n+1)u = 0, \qquad u'(1, n) + (k+1)u(1, n) = f(n);$$

hence

$$u(r, n) = f(n)r^n/(n + k + 1) \qquad (0 \leqq r \leqq 1, k > -1). \tag{36}$$

It follows that

$$r^{-k}(r^{k+1}u)' = f(n)r^n = v(r, n),$$

where $v(r, n)$ is the transform of the solution $V(r, x)$ of the Dirichlet problem (26), (27); therefore

$$u(r, n) = \int_0^r v(s, n)(s/r)^k \, ds/r = \int_0^1 v(rt, n)t^k \, dt.$$

The inverse transformation then gives the formula

$$U(r, \cos\theta) = \int_0^1 V(rt, \cos\theta)\, t^k \, dt \tag{37}$$

for the solution of problem (35) of the third type in terms of the solution of the Dirichlet problem (26), (27) with the same function $F(x)$.

A generalization of formula (37) can be verified easily. Let the above functions $U$, $V$ and $F$ be replaced here by $U(r, \phi, \theta)$, $V(r, \phi, \theta)$ and $F(\phi, \theta)$. For the sake of simplicity assume that $F$ is a continuous function of its two variables.

THEOREM 6: The potential function $U(r, \phi, \theta)$ of the third boundary value problem for the sphere,

$$\nabla^2 U = 0 \ (r < 1), \qquad U_r(1, \phi, \theta) + (k+1)U(1, \phi, \theta) = F(\phi, \theta) \tag{38}$$

where $k > -1$ and constant and the function $F$ is continuous, is given in terms of the solution $V(r, \phi, \theta)$ of the Dirichlet problem

$$\nabla^2 V = 0 \ (r < 1), \qquad V(1, \phi, \theta) = F(\phi, \theta) \tag{39}$$

by the formula

$$U(r, \phi, \theta) = \int_0^1 V(rt, \phi, \theta) t^k \, dt. \tag{40}$$

The verification that the function $U$ given by formula (40) is harmonic in the region $r < 1$ when $V$ is harmonic there is straightforward. To show that the function $U$ satisfies the boundary condition in problem (38) let $rt = s$ and note that

$$rU_r(r, \phi, \theta) = \int_0^1 V_s(s, \phi, \theta) t^{k+1} \, r dt = \int_0^1 Vt(s, \phi, \theta) t^{k+1} \, dt$$

when $r < 1$. Hence

$$\begin{aligned} &rU_r(r, \phi, \theta) + (k+1)U(r, \phi, \theta) \\ &= \int_0^1 [t^{k+1} V_t(s, \phi, \theta) + (k+1) t^k V(s, \phi, \theta)] \, dt = V(r, \phi, \theta) (r < 1). \end{aligned} \tag{41}$$

It is known that the function $V$ is continuous in the region $r \leqq 1$ and it follows from formula (40) that $U$ is continuous there. According to equation (41) then $rU_r$ is also continuous there. When $r = 1$ equation (41) reduces to the boundary condition in problem (38).

A Poisson integral formula for the solution of the third boundary value problem (38) follows at once by substituting the known Poisson integral that represents the solution $V$ of the Dirichlet problem into formula (40). Other forms of the Poisson integral formula for the special case (35) in which $U$ is independent of $\phi$ can be found from the formula (36) for $u(r, n)$ with the aid of the convolution property. For the inverse transform of $r^n(n + k + 1)^{-1}$ can be found either from transforms No. 10 and No. 6 in the table or by integrating the transformation No. 4 after multiplying by $a^k$.

The relation (40) between the solutions of the two types of problems for the sphere is so simple that it should be known. Neither that relation nor a Poisson integral formula for the problem of the third kind for the sphere nor the corresponding results in logarithmic potential for the circle, which follow similarly from finite Fourier transformations, have been found yet in the literature.

If $Y(r, \cos \theta)$ denotes the potential in the exterior region $r > 1$ for the third problem

$$\nabla^2 Y = 0 \ (r > 1), \qquad Y_r(1, x) - kY(1, x) = -F(x) \tag{42}$$

where $k > 0$ here and constant and $x = \cos \theta$, the transform of $Y$ is easily found to be

$$y(r, n) = (1/r) f(n) r^{-n} / (n + k + 1) = (1/r) u((1/r), n) \qquad (r \geqq 1)$$

where $u(r, n)$ is the transform (36) of the solution of the third problem (35) for the interior of the sphere. Hence

$$Y(r, \cos \theta) = (1/r) U(1/r, \cos \theta) \qquad (r > 1). \tag{43}$$

Moreover, the procedure used to arrive at formula (37) is easily applied to the above formula for $y(r, n)$ to arrive at the formula

$$Y(r, \cos\theta) = \int_1^\infty W(rt, \cos\theta) t^{-\frac{1}{2}-1}\, dt \tag{44}$$

where $W(r, \cos\theta)$ is the potential (29) in the Dirichlet problem for the exterior region,

$$\nabla^2 W = 0 \ (r > 1), \qquad W(1, \cos\theta) = F(\cos\theta).$$

The verification of formulas (43) and (44) shows that the potentials and the function $F$ involved need not be independent of the cooridnate $\phi$.

**7. Distributed sources in sphere.** As a final illustrative application of the transforms consider the problem

$$\nabla^2 V(r, x) = F(r, x) \ (r < 1), \qquad V(1, x) = 0 \tag{45}$$

where $x = \cos\theta$. The transformed problem is

$$r^2 v''(r, n) + 2rv'(r, n) - n(n+1)v(r, n) = r^2 f(r, n), \qquad v(1, n) = 0.$$

Its solution can be written by elementary methods in the form

$$v(r, n) = \int_0^1 \frac{(rs)^n}{2n+1} f(s, n) s^2\, ds - \frac{1}{r}\int_0^r \frac{s^n r^{-n}}{2n+1} f(s, n) s^2\, ds - \int_r^1 \frac{r^n s^{-n}}{2n+1} f(s, n) s\, ds. \tag{46}$$

According to transform No. 6 in the table,

$$T\{G(a, x)\} = a^n/(2n+1) \qquad (|a| \leqq 1)$$

where

$$G(a, x) = (1/2)(1 + a^2 - 2ax)^{-\frac{1}{2}}.$$

If $H(a, s, x)$ denotes the convolution (Theorem 3) of the two functions $G(a, x)$ and $F(s, x)$ as functions of $x$ the formal inverse transformation of the members of equation (46) gives the formula

$$V(r, x) = \int_0^1 H(rs, s, x) s^2\, ds - \frac{1}{r}\int_0^r H\left(\frac{s}{r}, s, x\right) s^2\, ds - \int_r^1 H\left(\frac{r}{s}, s, x\right) s\, ds$$

for the solution of problem (45).

## REFERENCES

1. Tranter, J. C., *Integral Transforms in Mathematical Physics*, Wiley, New York, 1951.
2. Sneddon, I. N., *Fourier Transforms*, McGraw-Hill, New York, 1951.
3. Burgat, P., "Résolution de problèmes aux limites au moyen de transformations fonctionnelles", *J. de Math. Phys. Appl.* **4**, 146–152 (1953).
4. Scott, E. J., "Jacobi Transforms", *Quart. J. Math. (Oxford)* **4**, 36–40 (1953).

5. CHURCHILL, R. V., "Integral Transforms and Boundary Value Problems", *Amer. Math. Monthly* **59**, 149–155 (1952).
6. CHURCHILL, R. V., AND DOLPH, C. L., "Inverse Transforms of Products of Legendre Transforms", *Proc. Amer. Math. Soc.* **5**, 93–100 (1954).
7. CHURCHILL, R. V., *Fourier Series and Boundary Value Problems*, McGraw-Hill, New York, 1941; 188.
8. KELLOGG, O. D., *Potential Theory*, Springer, Berlin, 1929.

THE UNIVERSITY OF MICHIGAN

(Received September 8, 1953)

# ASYMPTOTIC PROPERTIES OF BESSEL FUNCTIONS AND THE RADIATION FROM A SYNCHROTRON

By David Park

**1. The radiation.** The emission of electromagnetic radiation by an electron moving in a circular path (due to the action of a magnetic field) has received a careful mathematical discussion,[1,2] and it has been shown that if the angular frequency of the electron in its orbit, $\omega_0$, is taken as the fundamental, the total power radiated in the $n^{\text{th}}$ harmonic of this frequency is

$$P_n = n\omega_0 \frac{e^2}{R}\left[2\beta^2 J'_{2n}(2n\beta) - (1-\beta^2)\int_0^{2n\beta} J_{2n}(x)\,dx\right] \tag{1}$$

where $e$ is the electron's charge, $R$ is the radius of its orbit, and $\beta$ is its velocity expressed as a fraction of the speed of light. Schwinger has investigated the asymptotic behavior of this formula as $n$ becomes very large and $\beta$ approaches unity. It follows quite simply from the asymptotic properties of Bessel functions that for $n$ large, but still smaller than a critical $n_c$ in the neighborhood of $(1-\beta^2)^{-3/2}$, $P_n$ increases with $n$ according to

$$(P_n)_{ca} \sim \frac{3^{1/6}}{\pi}\,\Gamma(\tfrac{2}{3})\,\omega_0 \frac{e^2}{R} n^{1/3} \tag{2}$$

and that for larger $n$ it drops sharply to small values. A remarkable feature of this result is that for existing high-energy accelerators $n_c$ approaches $10^7$, and some of the radiation is visible to the eye.[3]

For values of $n$ on both sides of $n_c$, Schwinger has given the approximation, related to Nicholson's formulas,[4] that

$$P_n \sim \frac{3^{3/2}}{4\pi}\,\omega_0 \frac{e^2}{R}\frac{1}{1-\beta^2}\frac{n}{n_c^2}\int_{n/n_c}^{\infty} K_{5/3}(x)\,dx$$

where $K$ is the Bessel function as defined by Watson and $n_c = \frac{3}{2}(1-\beta^2)^{-3/2}$. The integral has been evaluated numerically[5] and the result found to agree well with experiment. The mathematical procedure has two disadvantages, however: that the accuracy of approximations of this kind is unknown, and that an infinite numerical integration is required.

The purpose of this paper is to investigate the behavior of the function $J_\nu(x)$ in the critical region of $x$ in the neighborhood of $\nu$ when $\nu$ is large, and to use the resulting formulas in order to evaluate (1) with an accuracy which can be estimated and in a form suitable for computation.

[1] G. A. Schott, *Phil. Mag.* (6) **13**, 189 (1907); *Electromagnetic Radiation* (Cambridge University Press, Cambridge, 1912).

[2] J. Schwinger, *Phys. Rev.* **75**, 1912 (1949).

[3] Elder, Gurewitsch, Langmuir, and Pollock, *Phys. Rev.* **71**, 829 (1947).

[4] G. N. Watson, *Bessel Functions* (Cambridge University Press, Cambridge, Second Edition, 1944).

[5] Elder, Langmuir, and Pollock, *Phys. Rev.* **74**, 52 (1948).

**2. The Behavior of $J_\nu(x)$ in the critical region.** The question of evaluating $J_\nu(x)$ for large real values of $\nu$ and $x$ in the neighborhood of $\nu$ has been studied in detail,[6] but the resulting formulas are not only unwieldy, but in some cases it is not even known to what extent they are accurate. We shall therefore approach the question from a different direction, and try to expand the function $J_\nu(\nu + \epsilon)$ in powers of $\epsilon$. For this we need the asymptotic values of $J_\nu$ and its derivatives at the point $\nu$. To find them we start with the formulas[7]

$$J_\nu(\nu) \sim \frac{\Gamma(\tfrac{1}{3})}{2^{2/3}3^{1/6}\pi\nu^{1/3}} - \frac{3^{7/6}\Gamma(\tfrac{5}{3})}{2^{1/3}\cdot 140\pi\nu^{5/3}} \tag{3}$$

$$J'_\nu(\nu) \sim \frac{3^{1/6}\Gamma(\tfrac{2}{3})}{2^{1/3}\pi\nu^{2/3}} + \frac{2^{1/3}\Gamma(\tfrac{4}{3})}{5\cdot 3^{1/6}\pi\nu^{4/3}} \tag{4}$$

and derive from them the higher derivatives by the use of Bessel's equation,

$$J''_\nu(Z) + \frac{1}{z}J'_\nu(z) + \left(1 - \frac{\nu^2}{z^2}\right)J_\nu(z).$$

Differentiating it $n$ times gives

$$J_\nu^{(2+n)}(\nu) + \sum_r r!\binom{n}{r}\binom{-1}{r}\nu^{-1-r}J_\nu^{(n-r+1)}(\nu) + J_\nu^{(n)}(\nu) - \sum_r r!\binom{n}{r}\binom{-2}{r}\nu^{-r}J_\nu^{(n-r)}(\nu) = 0$$

of which the dominant terms as $\nu$ becomes large are

$$J_\nu^{(2+n)}(\nu) \sim -\frac{1}{\nu}\,[J_\nu^{(n+1)}(\nu) + 2nJ_\nu^{(n-1)}(\nu)]. \tag{5}$$

Neglecting all but the first terms of (3) and (4), which are already accurate to within a few percent when $\nu = 10$, we can write

$$J'_\nu(\nu) \sim C\nu^{-1/3}J_\nu(\nu), \qquad C = 0.919\cdots \tag{6}$$

and then the first few terms of Taylor's series are found to be

$$\begin{aligned} J_\nu(\nu+\epsilon) = J_\nu(\nu)\Big[1 &+ \epsilon C\nu^{-1/3} - \frac{\epsilon^2}{2!}C\nu^{-4/3} - \frac{\epsilon^3}{3!}2\nu^{-1} \\ &- \frac{\epsilon^4}{4!}4C\nu^{-4/3} + \frac{\epsilon^5}{5!}10C\nu^{-7/3} + \frac{\epsilon^6}{6!}16\nu^{-2} \\ &+ \frac{\epsilon^7}{7!}40C\nu^{-7/3} - \frac{\epsilon^8}{8!}160C\nu^{-10/3} - \frac{\epsilon^9}{9!}224\nu^{-3} \\ &- \frac{\epsilon^{10}}{10!}640C\nu^{-10/3} + \cdots\Big]. \end{aligned} \tag{7}$$

[6] Ref. 4, Chapter VIII.

[7] Ref. 4, pp. 232, 247. (Note that in (2), p. 232, $3^{5/6}$ should be $3^{7/6}$.)

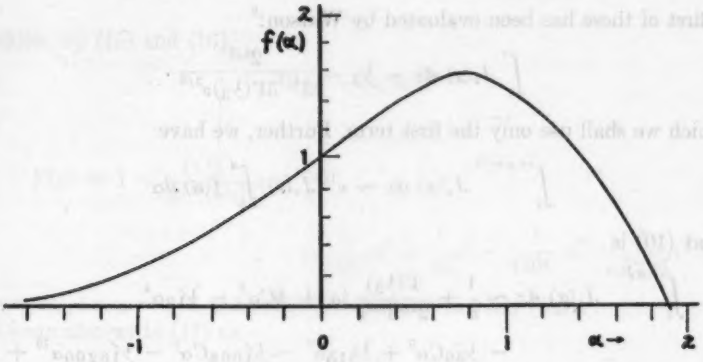



FIG. 1

It is found also that the exponent of $\nu$ springing from the iteration of (5) decreases by unity for every three steps in $n$. Setting $\epsilon$ equal to $\alpha\nu^{1/3}$ in (7) and omitting some terms proportional to $\nu^{-2/3}$ gives

$$J_\nu(\nu + \alpha\nu^{1/3}) \sim J_\nu(\nu)[1 + C\alpha - \tfrac{1}{3}\alpha^3 - \tfrac{1}{6}C\alpha^4 + \tfrac{1}{45}\alpha^6 + \tfrac{1}{126}C\alpha^7 - \tfrac{1}{1620}\alpha^9 - \tfrac{1}{5760}C\alpha^{10} + \cdots] \equiv J_\nu(\nu)f(\alpha) \tag{8}$$

from which we have the

THEOREM: *The quantity*

$$\frac{J_\nu(\nu + \alpha\nu^{1/3})}{J_\nu(\nu)}$$

*is, asymptotically, a function of* $\alpha$ *alone.* This function is plotted as $f(\alpha)$ in Fig. 1. The theorem implies that, asymptotically, the first zero of $J_\nu(x)$ occurs at $x = \nu + \alpha_1\nu^{1/3}$ for some fixed $\alpha_1$ independent of $\nu$, and similarly for the first maximum and subsequent zeros and maxima.[8]

We shall also need the corresponding formula for the derivative:

$$J'_\nu(\nu + \alpha\nu^{1/3}) \sim J'_\nu(\nu)\left[1 - \frac{1}{C}\alpha^2 - \frac{2}{3}\alpha^3 + \frac{2}{15C}\alpha^5 + \frac{1}{18}\alpha^6 - \frac{1}{180C}\alpha^8 - \frac{1}{567}\alpha^9 + \cdots\right]$$

and that for the integral, which we shall write as

$$\int_0^{\nu+\alpha\nu^{1/3}} J_\nu(x)\,dx = \int_0^{\nu} J_\nu(x)\,dx + \int_\nu^{\nu+\alpha\nu^{1/3}} J_\nu(x)\,dx. \tag{10}$$

[8] The formulas for the first zeros and maxima were given by Airey, *Phil. Mag.* (6) **34**, 193 (1917); they are discussed by Watson, Ref. 4, pp. 516, 521. The series (8) gives $\alpha_1 = 0.809$, Airey gives 0.763, and Watson gives 0.808618. We are also in agreement with Watson, as well as with Airey, on the position of the first zero.

The first of these has been evaluated by Watson:[9]

$$\int_0^{\nu} J_\nu(x)\,dx \sim \tfrac{1}{3} - \frac{2^{2/3}}{3^{1/3}5\Gamma(\tfrac{1}{3})\nu^{2/3}} \tag{11}$$

of which we shall use only the first term. Further, we have

$$\int_{\nu}^{\nu+\alpha\nu^{1/3}} J_\nu(x)\,dx \sim \nu^{1/3}J_\nu(\nu)\int_0^{\alpha} f(\alpha)\,d\alpha \tag{12}$$

so that (10) is

$$\int_0^{\nu+\alpha\nu^{1/3}} J_\nu(x)\,dx \sim \frac{1}{3} + \frac{\Gamma(\tfrac{1}{3})}{2^{2/3}3^{1/3}\pi}\left[\alpha + \tfrac{1}{2}C\alpha^2 - \tfrac{1}{12}\alpha^4 - \tfrac{1}{30}C\alpha^5 + \tfrac{1}{315}\alpha^7 - \tfrac{1}{1008}C\alpha^8 - \tfrac{1}{16200}\alpha^{10} + \cdots\right]. \tag{13}$$

As regards the accuracy of the approximation method, it is possible to find the next terms in the expressions of which we have given the leading ones by the use of the subsequent terms of (3), (4), (5), and (11), but they would contribute quite negligibly to results which we are seeking.

**3. The radiation spectrum.** We can now return to (1) and evaluate it as a power series in $\alpha$. First, however, we can note some of its general aspects. We shall suppose that $\beta$ is close to unity, and shall write

$$E = E_0(1 - \beta^2)^{-\frac{1}{2}} \tag{14}$$

for the electron's relativistic energy in terms of its rest energy. Now for a given $n$ and $E$ ($\gg E_0$), we can find the corresponding value of the parameter $\alpha$:

$$2n\beta = 2n + \alpha(2n)^{1/3}$$

$$\alpha = -(2n)^{2/3}(1 - \beta) \approx -(2n)^{2/3}(1 - \beta)\,\frac{1 + \beta}{2}$$

or

$$\alpha = -2^{-1/3}n^{2/3}(E_0/E)^2. \tag{15}$$

It is clear that $\alpha$ remains close to zero as $n$ increases until it reaches a value in the neighborhood of

$$n_0 = 2^{1/2}(E/E_0)^3 = 2^{3/2}\cdot 3n_c \tag{16}$$

at which $\alpha = -1$. Below $n_0$ we can use the asymptotic forms (4) and (11) in (1), and $P_n$ increases with $n$ approximately in accordance with (2). Above $n_0$, $P_n$ begins to decrease. The final formula for $P_n$ is found by substituting the above results into (1). It is

$$P_n \sim (P_n)_{cs}F(\alpha) \tag{17}$$

[9] G. N. Watson, *Phil. Mag.* (6) **35**, 364 (1918). (Note that $n^{2/3}$ should be $n^{1/3}$ in the relevant formula on p. 365.) See also Ref. 4, p. 259.

in which, by (15) and (16),

$$\alpha = -(n/n_0)^{2/3} \tag{18}$$

and

$$F(\alpha) = 1 + \frac{\Gamma(\tfrac{2}{3})}{2^{1/3}3^{1/3}C}\alpha - \frac{1}{6}\alpha^3 + \frac{1}{20C}\alpha^5 + \frac{1}{45}\alpha^6 - \frac{1}{420C}\alpha^8 - \frac{1}{1296}\alpha^9 + \cdots. \tag{19}$$

At $n = n_0$, i.e. $\alpha = -1$, the power has dropped to 0.321 $(P_n)_{\max}$.

We can also write (17) as

$$P_n \sim \frac{3^{1/6}}{\pi}\Gamma(\tfrac{2}{3})\,\omega_0\frac{e^2}{R}n_0^{1/3}(-\alpha)^{\frac{1}{2}}F(\alpha) \tag{20}$$

or, by (16),

$$P_n \sim \frac{6^{1/6}}{\pi}\Gamma(\tfrac{2}{3})\,\omega_0\frac{e^2}{R}\frac{E}{E_0}(-\alpha)^{\frac{1}{2}}F(\alpha). \tag{21}$$

This increases with $\alpha$ until it reaches a maximum at $\alpha = 0.41$, corresponding to $n = 0.37\ (E/E_0)^2$. The function $(-\alpha)^{1/2}F(\alpha)$ is plotted in Fig. 2. It is remarkable that the harmonic in which the radiation is greatest is a function only of the electron's velocity in its orbit. The General Electric synchrotron on which Elder, Langmuir, and Pollock[3] took their measurements has a fundamental frequency of 163.5 mc, and our formula predicts that, leaving out the fact that the electron is being accelerated as it goes, the peak of the spectrum will occur at a wavelength

$$\lambda = \frac{6.6\cdot 10^9}{\mathcal{E}^2}\,\text{Å}$$

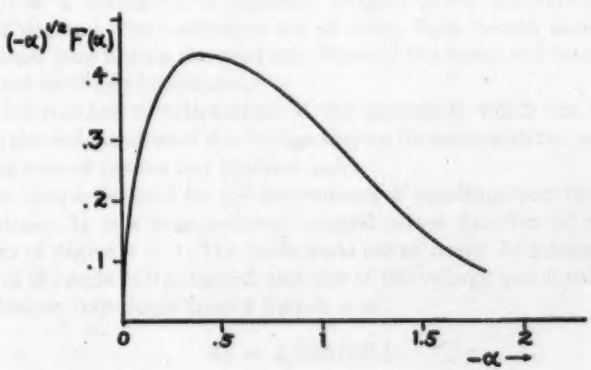



FIG. 2

where $\mathcal{E}$ is the electron's energy in mev. These results correspond well with those obtained previously[5] by numerical integration of Schwinger's approximate formula.

This work was done five years ago at the University of Michigan, assisted by the grant of a Rackham fellowship, for which the author offers his belated thanks.

*Note added in proof:* Since submitting this paper we have noticed that the series in (7) can be expressed in terms of a tabulated function, for if we substitute (8) into Bessel's equation and take $|\alpha| \ll \nu^{2/3}$ we find that $f(\alpha)$ satisfies

$$f''(\alpha) + 2\alpha f(\alpha) = 0$$

which is a form of Airy's equation. The solution of this corresponding to the series (8) is

$$f(\alpha) = 3^{2/3}\Gamma(\tfrac{2}{3})\,\mathrm{Ai}\,(-2^{1/3}\alpha)$$

where $\mathrm{Ai}(x)$ is Airy's function as defined by Jeffreys.[10] This result clarifies the relation of our formulas to those of Nicolson[4] and Tricomi.[11]

WILLIAMS COLLEGE

(Received September 8, 1953)

[10] H. and B. Jeffreys, *Methods of Mathematical Physics* (Cambridge University Press, Cambridge, second edition, 1950).

[11] F. Tricomi, *Atti Acad. Sci. Torino* **83**, 1 (1948). I am indebted to Dr. A. Erdélyi for calling this paper to my attention.

# ON ELECTRICAL NETWORK DETERMINANTS

By N. F. Tsang

In dealing with problems in electrical networks, Kirchhoff's Laws (which are in fact existence theorems of current and potential) are usually taken as a starting point. Then the system of loop equations or node equations is developed, whichever is more convenient for the problem on hand.

Maxwell has been usually accredited with the loop equations, although Kirchhoff probably had obtained them earlier[1]. On the other hand, node equations were systematically developed in Maxwell's classical treatise on Electricity and Magnetism. Loop equations have been more convenient in most cases and therefore very much more widely adopted up until the development of vacuum tube circuits. In general the two systems of equations are both linear simultaneous algebraic, but the number of unknowns, or order, are different. The order of loop system equations is equal to the number of independent loops $l$, while that of the node system, one less than the number of junction points or nodes, $n - 1$. Thus, the loop equations are to be preferred if $l < n - 1$, and vice versa.

As the two procedures yield different number of equations, they are sometimes considered as quite unrelated, and frequent claims have been made on the relative merits of one over the other. Although the two determinants are of different order, the loop determinant $\Delta$ being $l \times l$ square, and the node determinant $\Delta'$, $n - 1 \times n - 1$ square, it will be shown in this paper that under certain conditions, they will have simple relations with each other.

Let a network be given with $l$ loops, $n$ nodes, and $k$ independent branches (that is, not coupled with any other) of impedances. Then the following facts are known:

1. The determinant $\Delta$ resulting from the system of loop equations will consist of a number of terms, each being a product of $l$ of the $k$ branch impedances $Z_\alpha$. It is therefore a homogeneous algebraic integral power function of the impedances of degree $l$. The coefficients are all unity. Each branch impedance will appear at least once among the products. None of the terms will involve square or higher powers of any impedance.

2. The determinant is independent of the manner in which the $l$ loops are taken. It is also independent of the voltage sources (in series with the impedances) and current sources (across any junction pair).

3. Similar properties hold for the determinant $\Delta'$ resulting from the system of node equations. It is a homogeneous integral power function of the branch admittances of degree $n - 1$. The coefficients are all unity. It is independent of the choice of the node as 0 potential, and also of the voltage and current sources.

4. The look-in impedance from a branch $\alpha$ is

$$Z'_\alpha = \Delta/(\partial\Delta/\partial Z_\alpha) \qquad (1)$$

5. The look-in admittance from the junction points of the terminals of branch $\alpha$ is

$$Y'_\alpha = \Delta'/(\partial\Delta'/\partial Y_\alpha) \tag{2}$$

6. The following relation holds in general:

$$k = l + n - 1 \tag{3}$$

It is well known that if $\Delta$ is a homogeneous function of the $Z$s (the apostrophe between $Z$ and s is omitted to avoid possible confusion with the dashed quantities) of degree $l$, then

$$\sum_{\alpha=1}^{k} Z_\alpha(\partial\Delta/\partial Z_\alpha) \tag{4}$$

Hence

$$\sum_{1}^{k} (Z_\alpha/Z'_\alpha) = l \tag{5}$$

Similarly,

$$\sum_{1}^{k} (Y_\alpha/Y'_\alpha) = n - 1 \tag{6}$$

Note that although $Y_\alpha$ is the reciprocal of $Z_\alpha$, $Y'_\alpha$ is not the reciprocal of $Z'_\alpha$. The symbols $Z'_\alpha$ and $Y'_\alpha$ will appear only once more in this paper, so no confusion should arise as a result of this notation.

THEOREM 1. The sum of the complex ratios of a branch impedance to the corresponding look-in impedance over all branches in a network (without coupling between impedances) is equal to the number of independent loops. The sum of the complex ratios of a branch admittance to the corresponding look-in admittance over all branches in a network is equal to the number of junction points less one.

A simpler case will now be considered in detail. Let each branch contain only a resistance and an inductance in series, so that

$$Z_\alpha = R_\alpha + L_\alpha p.$$

Here $p$ is used instead of the conventional $j\omega$, because it is intended to study the behavior of the network toward any voltage or current of the type $e^{pt}$ (in this case $p$ will be real and negative). $\Delta$ will be a polynomial of $p$ of $l^{\text{th}}$ degree, so it will have $l$ roots. Let the roots of $\Delta$ and those of the $Z$s be subjected to the following conditions:

a. All the roots of $\Delta$ are different from each other.

b. All the roots of $\Delta$ are different from any of those of the $Z$s. The root of $Z_\alpha$, for example, is $-R_\alpha/L_\alpha$.

c. All the roots of $\Delta$ are different from any of those of the $\partial\Delta/\partial Z$s.

There will be no root for $\Delta'$ and the $Y$s except the trivial value of infinity.

The above 3 conditions assure that the look-in impedance from any branch will have the same set of $l$ roots. In general the above three conditions are usually satisfied if $R$s and $L$s are taken at random. Condition c is violated, e.g.,

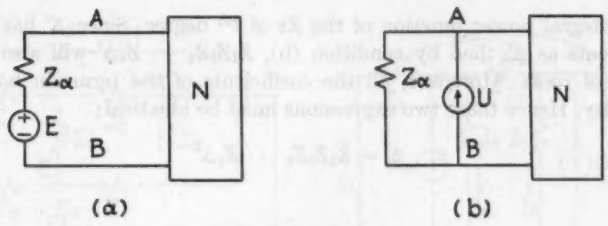



FIG. 1

by a balanced bridge circuit. If all the above three conditions are satisfied, the network is called non-degenerate.

Let the terminals of one of the impedances, $Z_\alpha$, be denoted by $A$, $B$. The loop look-in impedance will be that appearing to the voltage source $E$, in Fig. 1a. Let the network be subjected to the previous conditions and the roots of the determinant be $p_1, p_2, p_3, \cdots p_l$. For $p$ equal to any of the roots, say $p_1$, $\Delta = 0$, and the look-in impedances is also 0, by condition c. By condition b, $Z_\alpha \neq 0$. Hence the look-in impedance of the rest of the network $N$ will be $-Z_\alpha \neq 0$.

The look-in admittance from the nodes $A$, $B$ is that appearing to the current source $U$, Fig. 1b. It is contributed by branch $\alpha$ and $N$ in parallel, and therefore is equal to the sum of the corresponding admittances. When $p = p_1$,

$$Y' = (1/Z_\alpha) + (1/-Z_\alpha) = 0,$$

since $Z_\alpha \neq 0$. (If $Z_\alpha = 0$, then the look-in admittance will be infinite.)

The converse is also true. For, if the look-in admittance from $AB$ is 0, then the partial network $N$ will necessarily appear to be an impedance of $-Z_\alpha$, since $Z_\alpha = 0$. Hence,

THEOREM 2. In a non-degenerative network each root of $\Delta$ will also be a root of $\Delta'$.

Now examine the form of $\Delta'$. It is a homogeneous integral power polynomial in $Y$s with no square in any term and each $Y$ appears at least once among the various products. In terms of the $Z$s it is a homogeneous power function of degree $-n + 1$.

When $\Delta'$ is multiplied by the product $Z_1Z_2Z_3 \cdots Z_k$, it will become a homogeneous integral power function of the $Z$s with the degree

$$-n + 1 + k = l.$$

The $Z$s are sufficient to offset the $Y$s in all the terms, since no term in $\Delta'$ contains square or higher powers of any $Y$. Hence the product $Z_1Z_2Z_3 \cdots Z_k$ is sufficient to convert $\Delta'$ into a homogeneous integral power function of the $Z$s.

Since each $Y$ appears at least once in some product, all the $Z$s are necessary for multiplication to convert $\Delta'$ into an integral power function of $Z$s. For if $Z_1$ is left out, then $Y_1$ will appear somewhere among the terms.

Hence the product $Z_1Z_2Z_3 \cdots Z_k$ is necessary and sufficient to convert $\Delta'$

into an integral power function of the $Z$s of $l^{\text{th}}$ degree. Since $\Delta'$ has the same set of $l$ roots as $\Delta$, then by condition (b), $Z_1Z_2Z_3 \cdots Z_k\Delta'$ will also have the same set of roots. Moreover, all the coefficients of the terms in both $\Delta$ and $\Delta'$ are unity. Hence these two expressions must be identical:

$$\Delta = Z_1Z_2Z_3 \cdots Z_k\Delta' \tag{7}$$

or,

$$\Delta' = Y_1Y_2Y_3 \cdots Y_k\Delta \tag{8}$$

Since Eqs. 7, 8 are algebraic identities, they do not depend upon the particular form of the $Z$s. In other words, they will hold true for any network, with no coupling between the branch impedances, disregarding the three conditions.

THEOREM 3. For a network with independent impedances the ratio of the determinant of the system of loop equations to that of the node equations is the product of all the branch impedances.

Consider the more general case in which the $Z$s may be any combinations of impedance elements. Eqs. 7, 8 will still hold, as explained above. If one of the impedances, say $Z_1$, has an identical root as $\Delta$, (both simple) then

$$\Delta' = \Delta/(Z_1Z_2Z_3 \cdots Z_k)$$

will not have this root. A similar condition exists when some of the $Y$s have coincident roots as $\Delta'$.

Several examples will be given for illustrative purpose. The procedure of setting up the loop and node equations for many common networks are well known. To save space, it will be omitted in examples #1 and 2: only the derived determinants are given in these examples.

*Example 1.* Fig. 2 shows a non-planar network: $k = 9$, $l = 4$, $n = 6$. The following loops are chosen:

Loop 1: *ABED*, Loop 2: *BCFE*, Loop 3: *ADCB*, Loop 4: *ADEF*. The determinant for the mesh equations is (The $Z$s are omitted. Thus the first term $Z_1 + Z_3 + Z_4 + Z_6$ is written for shortness $1 + 3 + 4 + 6$)

$$\Delta = \begin{vmatrix} 1+3+4+6 & -4 & -1-3 & -3-6 \\ -4 & 2+4+5+7 & -2 & -7 \\ -1-3 & -2 & 1+2+3+8 & 3 \\ -3-6 & -7 & 3 & 3+6+7+9 \end{vmatrix}$$

$$\begin{aligned} = {} & 1235 + 1236 + 1237 + 1238 + 1256 + 1257 + 1259 + 1268 + 1269 + 1278 \\ & + 1279 + 1289 + 1345 + 1346 + 1347 + 1348 + 1356 + 1358 + 1367 + 1378 \\ & + 1456 + 1457 + 1459 + 1468 + 1469 + 1478 + 1479 + 1489 + 1567 + 1568 \\ & + 1569 + 1578 + 1589 + 1678 + 1679 + 1789 + 2345 + 2346 + 2347 + 2348 \\ & + 2357 + 2359 + 2367 + 2369 + 2378 + 2379 + 2389 + 2456 + 2457 + 2459 \\ & + 2468 + 2469 + 2478 + 2479 + 2489 + 2567 + 2569 + 2678 + 2679 + 2689 \\ & + 3456 + 3457 + 3458 + 3459 + 3469 + 3479 + 3489 + 3567 + 3569 + 3578 \\ & + 3589 + 3679 + 3789 + 4568 + 4578 + 4589 + 4689 + 4789 + 5678 + 5689 \\ & + 6789 \end{aligned}$$

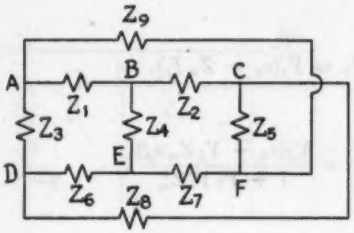



FIG. 2

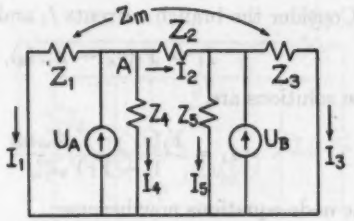



FIG. 3

For the node equations, take $D$ as 0 potential, and $ABCEF$ in order. The letter $Y$ will be omitted, thus: $Y_1 + Y_3 + Y_9$ will be written as $1 + 3 + 9$. The determinant $\Delta'$ will then be:

$$\Delta' = \begin{vmatrix} 1+3+9 & -1 & 0 & 0 & -9 \\ -1 & 1+2+4 & -2 & -4 & 0 \\ 0 & -2 & 2+5+8 & 0 & -5 \\ 0 & -4 & 0 & 4+6+7 & -7 \\ -9 & 0 & -5 & -7 & 5+7+9 \end{vmatrix}$$

$$\begin{aligned} = {} & 12345 + 12347 + 12349 + 12356 + 12357 + 12367 + 12369 + 12379 + 12456 \\ & + 12458 + 12467 + 12469 + 12478 + 12489 + 12567 + 12568 + 12578 + 12678 \\ & + 12679 + 12689 + 12789 + 13457 + 13458 + 13459 + 13478 + 13489 + 13567 \\ & + 13568 + 13569 + 13578 + 13579 + 13678 + 13689 + 13789 + 14567 + 14568 \\ & + 14569 + 14578 + 14589 + 14678 + 14689 + 15679 + 15689 + 15789 + 16789 \\ & + 23456 + 23458 + 23459 + 23467 + 23469 + 23478 + 23479 + 23489 + 23567 \\ & + 23568 + 23569 + 23578 + 23579 + 23678 + 23689 + 23789 + 24569 + 24589 \\ & + 24679 + 24789 + 25679 + 25689 + 25789 + 26789 + 34567 + 34568 + 34569 \\ & + 34578 + 34579 + 34678 + 34689 + 34789 + 45679 + 45689 + 54789 + 46789 \end{aligned}$$

Each determinant contains 81 terms. The relation given by Eqs. (7), (8) can be shown to hold by checking each term of $\Delta'$ against another in $\Delta$. For example, the first term of $\Delta'$, 12345 or $Y_1Y_2Y_3Y_4Y_5$, when multiplied by $Z_1Z_2Z_3 \cdots Z_9Z_9$ gives $Z_6Z_7Z_8Z_9$ or 6789, corresponding to the last term of $\Delta$.

It should be noted that Fig. 2 shows the simplest non-planar network. The two determinants obtained above are therefore simplest among those in connection with non-planar networks.

We shall now examine the effect of couplings between the various branch elements. These couplings may be in the form of mutual or transimpedance or transadmittance, bilateral or unilateral. In general, they may be called transimmittance[2]. Theorem 3 was derived from Theorem 2, and the fact that the product $Z_1Z_2Z_3 \cdots Z_k$ is necessary and sufficient to convert $\Delta'$ into a homogeneous integral power function of the $Z$s. Theorem 2 is quite general and independent of any transimmittance between the $Z$s. The latter relation, however, depends upon the algebraic form of $\Delta'$, and so is affected by these transimmittances.

Fig. 3 shows a 3-loop network with a bilateral mutual impedance $Z_m$ between $Z_1$ and $Z_3$. The loop equations can be easily derived. The coefficients of the circulating currents are all linear combinations of $Z_1$, $Z_2$, $Z_3$, $Z_4$, $Z_5$ and $Z_m$, so that $\Delta$ will be a homogeneous integral power function of the $Z$s, including $Z_m$.

Consider the branch currents $I_1$ and $I_3$ :

$$I_1 = Y_1(v_A - Z_m I_3), \qquad I_3 = Y_3(v_B - Z_m I_1).$$

The solutions are

$$I_1 = \frac{Y_1(v_A - Y_3 Z_m v_B)}{1 - Y_1 Y_3 Z_m^2} \qquad I_3 = \frac{Y_3(v_B - Y_1 Z_m v_A)}{1 - Y_1 Y_3 Z_m^2}.$$

The node equations now become:

$$\left(\frac{Y_1}{1 - Y_1 Y_3 Z_m^2} + Y_2 + Y_4\right) v_A - \left(Y_2 + \frac{Y_1 Z_m Y_3}{1 - Y_1 Y_3 Z_m^2}\right) v_B = U_A$$

$$-\left(Y_2 + \frac{Y_1 Y_3 Z_m}{1 - Y_1 Y_3 Z_m^2}\right) v_A + \left(\frac{Y_3}{1 - Y_1 Y_3 Z_m^2} + Y_2 + Y_4\right) v_B = U_B$$

Thus the determinant $\Delta'$ will contain the factor $1 - Y_1 Y_3 Z_m^2$ in the denominator: it will be evident if the coefficients for $Y_4$ are collected. In the present case, therefore, Eqs. (7), (8) do not hold.

Next consider the same case, but with $Z_m$ unilateral, from $Z_1$ to $Z_3$, that is, $I_1$ will cause a voltage drop in branch 3, but $I_3$ will not cause any drop in branch 1. Then

$$I_1 = Y_1 v_A, \qquad I_3 = Y_3(v_B - Z_m I_1) = Y_3(v_B - Z_m Y_1 v_A)$$

The node equations are:

$$(Y_1 + Y_2 + Y_4)v_A - Y_2 v_B = U_A$$

$$-(Y_2 + Y_1 Y_3 Z_m)v_A + (Y_2 + Y_3 + Y_5)v_B = U_B.$$

Hence the determinant $\Delta'$ will still be an integral algebraic power function of $Y$s and $Z_m$, although some terms will contain more $Y$s than the others, being compensated by $Z_m$. No square or higher powers are present. Multiplied by $Z_1 Z_2 Z_3 Z_4 Z_5$, it will become an integral algebraic homogeneous function in $Z_1, \cdots, Z_m$. By the same line of reasoning, we see that Eqs. (7) and (8) will still hold.

*Definition.* A simply connected set of unilateral transimmittances in a network consists of a number of transimmittances, the whole or any part of them do not form any closed loop. For example, a simply connected set would be formed by the following group of transimpedances: 1. branch 1 to 3, 2. branch 3 to 6, and 3. branch 1 to 6. If the third one is from 6 to 1, then 1–3–6–1 will form a closed loop, and the set will be no more simply connected.

The above generalization may be extended to networks with more than one unilateral transimmittances, provided they form a simply connected set.

THEOREM 4. For a network with impedances containing transimmittances forming a simply connected set, the ratio of the determinant of the system of loop equations to that of the node equations is the product of all the branch impedances. This ratio is independent of any of the transmimittances present.

The proof of the above theorem is necessarily very lengthy. Moreover, mathe-

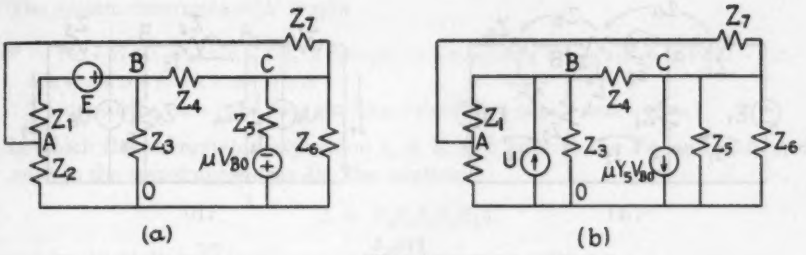



FIG. 4

matical rigor is hard to attain. Instead, two examples will be given as a check on the results. Example 2 involves the familiar vacuum tube transconductance. In example 3 a number of transimpedances are present, so the procedure of obtaining the loop and node equations is also given.

*Example 2.* Fig. 4 shows a simple feedback amplifier with 4 loops, 4 nodes, and 7 branches. The following loops are chosen:

Loop 1: $ABO$, Loop 2: $BCO$, Loop 3: $Z_5$ and $Z_6$, Loop 4: $ACB$. The determinant $\Delta$ is:

$$\Delta = \begin{vmatrix} 1+2+3 & -3 & 0 & -1 \\ -3-8 & 3+4+5+8 & -5 & -4 \\ 8 & -5-8 & 5+6 & 0 \\ -1 & -4 & 0 & 1+4+7 \end{vmatrix} \quad \text{where } Z_5 = \mu Z_3.$$

Expanding by the usual process, we obtain,

$$\begin{aligned} \Delta = {} & 1235 + 1236 + 1245 + 1246 + 1256 + 1268 + 1356 + 1357 + 1367 \\ & + 1456 + 1457 + 1467 + 1567 + 1678 + 2345 + 2346 + 2357 + 2367 \\ & + 2456 + 2457 + 2467 + 2468 + 2567 + 2678 + 3456 + 3457 + 3467 \\ & + 3567 \end{aligned}$$

The node equations are obtained by taking $ABC$ in order as nodes and point 0 as ground. The determinant becomes

$$\Delta' = \begin{vmatrix} 1+2+7 & -1 & -7 \\ -1 & 1+3+4 & -4 \\ -7 & 9-4 & 4+5+6+7 \end{vmatrix} \quad \text{where } Y_5 = \mu Y_3$$

$$\begin{aligned} = {} & 124 + 125 + 126 + 127 + 134 + 135 + 136 + 137 + 145 + 146 \\ & + 149 + 157 + 167 + 179 + 234 + 235 + 236 + 237 + 245 + 246 \\ & + 247 + 249 + 347 + 357 + 367 + 457 + 467 + 479. \end{aligned}$$

The validity of Eqs. (7) and (8) can again be demonstrated by checking term for term of $\Delta$ against $\Delta'$. The term 479 or $Y_4Y_7Y_9$ in $\Delta'$, when multiplied by $Z_1Z_2Z_3Z_4Z_5Z_6Z_7$, gives $Z_1Z_2Z_6Z_8$ or 1268, since $Z_3Z_5Y_9 = Z_3Z_5\mu Y_3 = \mu Z_3 = Z_5$.

*Example 3.* Fig. 5 shows a network with 5 branches, 3 junctions and 3 loops.

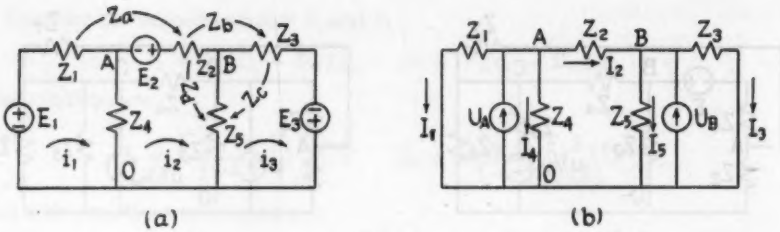



FIG. 5

There are 4 unilateral transimpedances simply connected. The loop equations may be written down as follows:

$$Z_1 i_1 + Z_4(i_1 - i_2) = E_1$$

$$Z_4(i_2 - i_1) + Z_2 i_2 + Z_a i_1 + Z_5(i_2 - i_3) + Z_c i_3 + Z_d i_2 = E_2$$

$$Z_5(i_3 - i_2) - Z_c i_3 - Z_d i_2 + Z_3 i_3 + Z_b i_2 = E_3 .$$

The system determinant is then

$$\Delta = \begin{vmatrix} 1+4 & -4 & 0 \\ a-4 & 2+4+5+d & c-5 \\ 0 & b-d-5 & 3+5-c \end{vmatrix}$$

$$= 123 + 125 - 12c + 134 + 135 + 13d + 145 - 14c + 15b - 1bc + 234 + 245 - 24c + 345 + 34a + 34d + 45a + 45b - 4ac - 4bc.$$

The setting up of the node equations are not so direct, but can be obtained from the following voltage drops:

$$Z_1 I_1 = v_A , \qquad Z_4 I_4 = v_A , \qquad Z_2 I_2 - Z_a I_1 = v_A - v_B ,$$

$$Z_3 I_3 + Z_b I_2 = v_B , \qquad Z_5 I_5 + Z_c I_3 + Z_d I_2 = v_B .$$

The currents are then found from the above equations:

$$I_1 = Y_1 v_A , \qquad I_2 = Y_2(1 + Y_1 Z_a) v_A - Y_2 v_B ,$$

$$I_3 = Y_3 - Y_3 Z_b(1 + Y_1 Z_a) v_A + (1 + Y_2 Z_b) v_B , \qquad I_4 = Y_4 v_A ,$$

$$I_5 = Y_5 Y_2(1 + Y_1 Z_a)(Y_3 Z_b Z_c - Z_d) v_A + (1 - Y_3 Z_c + Y_2 Z_d - Y_2 Y_3 Z_b Z_c) v_B .$$

If there are current sources $U_A$ and $U_B$ at $A$, $B$ respectively, then

$$(Y_1 + Y_2 + Y_4 + Y_1 Y_2 Z_a) v_A - Y_2 v_B = I_1 + I_2 + I_4 = U_A$$

$$Y_2(1 + Y_1 Z_a)(-1 - Y_3 Z_b - Y_5 Z_d + Y_3 Y_5 Z_b Z_c) v_A + Y_2 + Y_3 + Y_5(1 + Y_2 Z_d - Y_3 Z_c) + Y_2 Y_3 Z_b(1 - Y_5 Z_c) v_B = -I_3 + I_5 + I_5 = U_B .$$

The system determinant $\Delta'$ is now

$$\Delta' = \begin{vmatrix} Y_1 + Y_2 + Y_4 + Y_1Y_2Z_a, & -Y_2 \\ Y_3(1 + Y_1Z_a)(-1 - Y_2Z_b - Y_4Z_d + Y_2Y_4Z_bZ_d), & Y_2 + Y_3 + Y_4(1 + Y_2Z_d - Y_2Z_a) + Y_2Y_4Z_b(1 - Y_2Z_a) \end{vmatrix}$$

$$= 12 + 13 + 15 + 23 + 24 + 25 + 34 + 45$$
$$+ 123a + 123b + 125a + 125d - 135c + 234b + 235c + 245d - 345c - 1235ac - 1235bc - 2345bc,$$

in which the abbreviated expression 1, 2, 3, 4, 5 refer to the $Y$s, and $a$, $b$, $c$, $d$ refer to the transimpedances $Z$s. The relation

$$\Delta = Z_1Z_2Z_3Z_4Z_5\Delta'$$

can be shown to hold in this case by checking such term, e.g.,

$$Z_1Z_2Z_3Z_4Z_5Y_2Y_3Y_4Z_b = Z_1Z_5Z_b \text{ or } 15b.$$

In the above example, if $Z_d$ acts unilaterally in the other direction (from 5 to 2), then there will be simultaneous equations between $I_2$, $I_3$, and $I_5$, and $\Delta'$ would be no more an integral power function of $Y_1, \cdots, Y_5, Z_a, \cdots Z_d$. Eq. (7) will not be satisfied.

The writer acknowledges certain suggestions from Mr. T. T. Wu, of the University of Minnesota.

## REFERENCES

1. Y. H. Ku: Resume of Maxwell's and Kirchhoff's Rules for Network Analysis, Jour Franklin Inst. Vol. 253, No. 3, March 1952.
2. H. W. Bode: Network Analysis and Feedback Amplifier Design, p. 15, D. Van Nostrand, New York, 1945.

University of Arkansas

(Received September 8, 1953)

# CONVERSION OF A BRUNE CYCLE WITH AN IDEAL TRANSFORMER INTO A CYCLE WITHOUT AN IDEAL TRANSFORMER*

F. M. Reza

**1. Introduction.** O. Brune (1) has proved that any rational "positive real" function possesses a one terminal-pair network configuration containing a finite number of linear passive elements. Brune suggested the ladder synthesis for positive real functions, but this is generally confronted by mutual couplings, which are not desirable from a practical standpoint. Thus there has always been a great desire to eliminate ideal transformers from the picture. As steps towards this goal, H. W. Bode suggests resistance padding, and E. A. Guillemin (2) often uses some practical methods for eliminating the ideal transformer in special cases. The existence proof, along with a synthesis procedure without an ideal transformer, was first given by R. Bott and R. J. Duffin (3).

The object of this paper is to outline some results which have been found by the author:

By the continuation of the Brune synthesis procedure we have arrived at Bott and Duffin's result. This provides an alternative proof to that of Bott and Duffin for the existence of a network configuration without an ideal transformer for any positive real function.

Our method of approach suggests a more unified synthesis procedure: Follow the Brune synthesis procedure. If a Brune cycle with an ideal transformer is confronted, convert that cycle into a cycle without a transformer. Numerical values of the elements of the latter cycle are found in terms of elements of the corresponding Brune cycle.

**2. Existence proof.** In the Brune synthesis of a driving-point impedance $Z(s)$, when Re $Z(j\omega)$has a finite minimum along the axis of real frequencies, one is confronted with an ideal transformer. Let $Z(s)$ be such a function, whose minimum real part has already been removed. It is known (1, 4) that the removal of elements $L_1$, $L_2$, $L_3$ ($L_3$ assumed negative), and $c$ as shown in Figure 1 will lead to a positive real function $z$ which may or may not contain an ideal transformer in its Brune configuration. From Figure 1 it is seen that

$$\frac{1}{Z - L_1 s} = \frac{cs}{cL_2 s^2 + 1} + \frac{1}{L_3 s + z}$$

$$Z_1(s) = Z - L_1 s = \frac{(L_3 s + z)(cL_2 s^2 + 1)}{[c(L_3 + L_2)s^2 + 1] + csz} \tag{1}$$

It is known to those using the Brune synthesis that the function $(Z - L_1 s)$ has only one positive real zero, a fact which necessitates an ideal transformer

* This work was supported in part by the Signal Corps, the Air Materiel Command, and the Office of Naval Research.

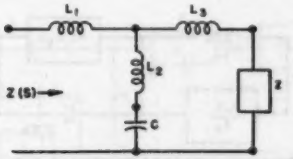



FIG. 1

in this synthesis. For the sake of future reference we may conveniently state this fact in the following theorem.

THEOREM: Let $Z(s)$ be a positive real function and $A$ a positive constant. Then $Z(s) - As = 0$ cannot have more than one zero in the right half-plane.

Proof of this theorem was first given by Brune (ref. 1, Theorem VI, Corollary I). An independent proof has also been established by the author. (This proof will not be presented in this paper.)

If the positive real root which enters Eq. 1 is denoted as $k$, then

$$Z(k) - L_1 k = 0 \tag{2}$$

$$z(k) + L_3 k = 0 \tag{3}$$

In an attempt to avoid the obstacle of the ideal transformer it seems natural at first to attempt the Brune synthesis of the reciprocal function on the admittance basis (see ref. 4, page 268).

$$[1/Z(s)] - As$$

where $A$ is a positive constant.

Based on the above theorem, it could be shown that if $Z(s)$ satisfies the conditions outlined above, then

$$Z_2 = Z^{-1} - As$$

has all its poles and zeros in the left half-plane, except for one positive real zero. One may select $A$ so that the positive real zero of $Z_2$ coincides with that of $Z_1$. In that case,

$$[1/Z(k)] - Ak = 0, \qquad A = 1/L_1 k^2 \tag{4}$$

Now consider the function

$$\frac{Z - L_1 s}{\frac{1}{Z} - (s/k^2 L_1)} = \frac{Z - L_1 s}{k^2 L_1 - sZ} \cdot k^2 L_1 Z$$

and let

$$Z_3(s) = k\left(\frac{Z - L_1 s}{k^2 L_1 - sZ}\right) \tag{5}$$

It can be shown that $Z_3$ is a positive real function: its poles and zeros are in

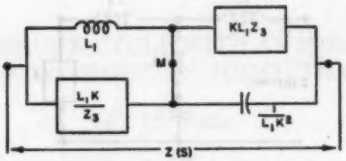



FIG. 2

the left half-plane, the difference of the degree of its numerator and denominator is not more than one, and Re $Z_3(j\omega) \geqq 0$, as shown below

$$\text{Re } Z_3(j\omega) = k \frac{L_1 U(k^2 + \omega^2)}{(k^2 L_1 + \omega V)^2 + \omega^2 U^2} \tag{6}$$

where

$$Z(j\omega) = U + jV$$

The network configuration of Figure 2 is an immediate consequence of Eqs. 5 and 7.

$$Z = \frac{1}{\dfrac{Z_3}{L_1 k} + \dfrac{1}{L_1 s}} + \frac{1}{\dfrac{s}{L_1 k^2} + \dfrac{1}{kL_1 Z_3}} \tag{7}$$

**3. Synthesis of $Z_3$.** Using formulas 1 and 5, we will find

$$\frac{1}{Z_3} = \frac{L_1(k^2 - s^2)[csz + c(L_2 + L_3)s^2 + 1] - s(L_3 s + z)(L_2 cs^2 + 1)}{k(L_3 s + z)(cL_2 s^2 + 1)} \tag{8}$$

which suggests that a pair of poles of $1/Z_3$ on the axis of imaginaries could be removed. To find the values of elements corresponding to this pair of poles, one can proceed by finding the proper residues:

$$\text{residue of } \left[\frac{1}{Z_3}\right] \text{ at } \frac{\pm j}{(L_2 c)^{\frac{1}{2}}} = \frac{L_1(k^2 L_2 c + 1)}{2kL_2^2 c} \tag{9}$$

Let the corresponding elements be $L_r$ and $c_r$. Then

$$L_r = \frac{L_1(k^2 L_2 c + 1)}{kL_2}$$

$$c_r = \frac{kL_2^2 c}{L_1(k^2 L_2 c + 1)} \tag{10}$$

The network of Figure 2 with proper impedance levels for $Z_3$ will lead, along with the results outlined in Eqs. 8 and 10, to the final conversion cycle such as that shown in Figure 3.

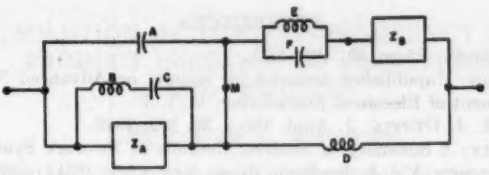



FIG. 3

In the particular case of a quadratic rational fraction, one obtains the equivalent network, without an ideal transformer, by letting $z = r$ in Figure 1, and

$$k = \frac{-r}{L_3}, \qquad Z_A = r, \quad \text{and} \quad Z_B = \frac{r(L_1 + L_2)}{L_2 + L_3} \tag{11}$$

in Figure 3.

It is important to note that the structure of Figure 3 and 2 represent a balanced bridge. In fact

$$L_1 s \times \frac{L_1 k^2}{s} = \frac{L_1 k}{Z_3} \times k L_1 Z_3 = k^2 L_1^2 \tag{12}$$

which shows that the opening of the node $M$ will not destroy the balance of the bridge. This reduces Figure 2 to one loop in conformity with Reference 5.

This method may be summarized as follows:

1. Proceed with the Brune synthesis in general. If confronted with an ideal transformer, finish the Brune cycle and find $L_1$, $L_2$, $L_3$, $c$ and $k$.
2. Draw Figure 3 and compute the proper values of $A$, $B$, $C$, $D$, $E$, $F$, $Z_3$, $Z_A$, and $Z_B$ as shown below.

$$A = 1/L_1 k^2 \qquad D = L_1$$

$$B = \frac{k^2 L_2^2 c}{k^2 L_2 c + 1} \qquad E = \frac{L_1^2(k^2 L_2 c + 1)}{L_2}$$

$$C = \frac{k^2 L_2 c + 1}{k^2 L_2} \qquad F = \frac{L_2^2 c}{L_1^2(k^2 L_2 c + 1)}$$

$$\frac{1}{Z_3} = \frac{L_1 k^2 - sZ}{k(Z - L_1 s)} = \frac{L_1(k^2 - s^2)[csz + c(L_2 + L_3)s^2 + 1]}{k(L_3 s + z)(cL_2 s^2 + 1)} - \frac{s}{k}$$

$$\frac{1}{Z_A} = \frac{1}{kL_1 Z_3} - \frac{(k^2 L_2 c + 1)s}{k^2 L_2(cL_2 s^2 + 1)} \qquad Z_B = \frac{L_1 k}{Z_3} - \frac{L_1^2(k^2 L_2 c + 1)s}{L_2(cL_2 s^2 + 1)}$$

3. Follow the Brune procedure on $Z_A$ and $Z_B$. If the Brune synthesis of $Z_A$ or $Z_B$ again introduces ideal transformers, proceed again as in step 2.

Further results are given in a Letter to the Editor, Journal of Applied Physics, 25, 807 (1954)

## REFERENCES

1. O. Brune: J. Math. Phys. **10**, 191, 1931.
2. E. A. Guillemin: Unpublished material for courses on Advanced Network Theory, Department of Electrical Engineering, M. I. T.
3. R. Bott and R. J. Duffin: J. Appl. Phys. **20**, 816, 1949.
4. E. A. Guillemin: A Summary of Modern Methods of Network Synthesis, Advances in Electronics, Vol. 3, Academic Press, New York, 1951, p. 268.
5. F. M. Reza: Quarterly Progress Report, Research Laboratory of Electronics, M. I. T. Jan. 15, 1953.

Research Laboratory of Electronics
Massachusetts Institute of Technology

(Received November 3, 1953)

# A GENERAL SOLUTION OF THE TWO-FREQUENCY MODULATION PRODUCT PROBLEM. III. RECTIFIERS AND LIMITERS

By Robert L. Sternberg

**1. Introduction.** In Part I of this paper by the present writer with H. Kaufman [3] it was shown that if a two-frequency input is applied to a continuous nonlinear device then the amplitudes of the modulation products which occur in the output can always be determined approximately as linear combinations of certain new functions $A_{mn}(h, k)$ introduced by W. R. Bennett [1], [2] and, moreover, that if the output versus input characteristic of the device is not only continuous but is polygonal as well, then the values of the modulation product amplitudes so determined are exact. In Part II of this paper by the present writer [4] short tables of the first ten functions $A_{mn}(h, k)$ were provided. The purpose of the present paper then is to apply the theory of Part I to the case of various ideal rectifiers and limiters having continuous polygonal output versus input characteristics to obtain exact expressions for the modulation product amplitudes in terms of the functions $A_{mn}(h, k)$ tabulated in Part II. Thus, in particular, using the notation of Part I, if the input to the ideal rectifier or limiter is given as

$$x(t) = P\cos(pt + \theta_p) + Q\cos(qt + \theta_q), \quad 0 < P < P + Q \leqq 2P, \tag{1.1}$$

so that the output is expressible in the form

$$y(t) = \tfrac{1}{2}B_{00} + \sum\nolimits^{*}_{m,n=0}\infty B_{\pm mn}\cos(\omega_{\pm mn}t + \phi_{\pm mn}), \tag{1.2}$$

where the asterisk has the same meaning as in (2.2) of Part I and

$$\omega_{\pm mn} = mp \pm nq, \qquad \phi_{\pm mn} = m\theta_p \pm n\theta_q, \qquad (m + n \geqq 1), \tag{1.3}$$

then the purpose of this paper is to obtain explicit formulas for the quantities $B_{\pm mn}$, $(m + n \geqq 0)$ in terms of the functions $A_{mn}(h, k)$. Numerical values of the modulation product amplitudes $\tfrac{1}{2}B_{00}$ and $B_{\pm mn}$, $(m + n \geqq 1)$ may then be obtained with the aid of the tables of Part II while a partial solution for the quantities $\tfrac{1}{2}B_{00}$ and $B_{\pm mn}$, $(m + n \geqq 1)$ of an analytic nature in the form of power series expansions of these quantities in powers of various associated parameters which converge in restricted regions will be indicated.

For purposes of application to actual non-ideal rectifiers and limiters we recall from Part I that if the output versus input characteristic of the actual device differs from that of the ideal device by not more than $\epsilon$ throughout the operating range then the values of the modulation product amplitudes in the output from the actual device will differ from the values obtained for the quantities $\tfrac{1}{2}B_{00}$ and $B_{\pm mn}$, $(m + n \geqq 1)$ by not more than $4\epsilon/\pi$ for all $m$ and $n$ while the output as given by (1.2) will differ from the actual output by not more than $\epsilon$ for all time.

The biased ideal rectifier is treated in §2, the full wave ideal rectifier is treated

in §3 and the symmetrical and unsymmetrical ideal limiters are treated in §§4 and 5. In §6 some closing remarks about terminology are made. The notation of Part I is used throughout. For purposes of future reference we note that the name "Bennett functions" has been proposed in Part II for the functions $A_{mn}(h, k)$.

**2. The biased ideal rectifier.** For the biased ideal rectifier with the output versus input characteristic

$$Y(X) = \begin{cases} 0, & X \leqq X_0, \\ g(X - X_0), & X \geqq X_0, \end{cases} \qquad g > 0, \tag{2.1}$$

$g$ being a given constant as indicated, we have by definition the result

$$B_{\pm mn} = PgA_{mn}(h, k), \qquad (m, n = 0, 1, 2, \cdots), \tag{2.2}$$

where $h = (X_0/P)$ and $k = (Q/P)$. Except for minor notational changes relations (2.2) are equivalent to the original results of Bennett [1], [2]. All of the properties of the functions $A_{mn}(h, k)$ given in Parts I and II apply. In particular, for analytical purposes we call attention to the double power series expansions of the four basic functions $A_{00}(h, k)$, $A_{10}(h, k)$, $A_{01}(h, k)$ and $A_{11}(h, k)$ for $|h| + k \leqq 1$ given in Part I, §5 which with the aid of the recursion identities of Part I, §4 can be combined to yield double power series expansions of the higher order functions $A_{mn}(h, k)$ for $|h| + k \leqq 1$ and for immediate numerical results we call attention to the tables of the first ten functions $A_{mn}(h, k)$ given in Part II. Finally, for the special case of the zero biased ideal rectifier we refer the reader to the complete solution for the $A_{mn}(0, k)$ in terms of tabulated complete elliptic integrals for $k \neq 1$, or rational functions of $\pi$ for $k = 1$, partially indicated in formulas (2.7) and (2.8) of Part II and given in detail in the original results of Bennett [1; pp. 234–235].

**3. The full wave ideal rectifier.** Consider now the full wave ideal rectifier having the output versus input characteristic

$$Y(X) = g|X|, \qquad g > 0, \tag{3.1}$$

$g$ again being a given constant as indicated. Applying the theory of Part I, §3 we see readily that in any operating range $-a \leqq X \leqq a$, $a \geqq 2P$, the full wave ideal rectifier characteristic (3.1) may be expressed exactly in the form (3.1) of Part I with, in particular,

$$X_1 = -a, \qquad X_2 = 0, \qquad g_1 = -g, \qquad g_2 = 2g, \qquad N = 2. \tag{3.2}$$

Hence, applying formulas (3.8) of Part I and simplifying with the aid of relations (4.3) and (4.8) of Part I, we have the results

$$\begin{aligned} B_{\pm mn} &= 2PgA_{mn}(0, k), && (m + n = 0, 2, 4, \cdots), \\ B_{\pm mn} &= 0, && (m + n = 1, 3, 5, \cdots), \end{aligned} \tag{3.3}$$

where again $k = (Q/P)$.

Except for the factor $Pg$ the coefficients (3.3) are functions of $k$ only and thus in this case taking $h = 0$ in the double power series expansions of the functions $A_{mn}(h, k)$ of Part I, §5 and applying the recursion identities of Part I, §4 we are led by (3.3) to the simple power series expansions

$$(3.4) \quad \tfrac{1}{2}B_{00} = (Pg/\pi)[2 + (\tfrac{1}{2})k^2 + (\tfrac{1}{32})k^4 + \cdots],$$

$$(3.5) \quad B_{20} = (Pg/3\pi)[4 - 3k^2 + (\tfrac{9}{16})k^4 + \cdots],$$

$$(3.6) \quad B_{11} = (Pgk/\pi)[2 - (\tfrac{1}{4})k^2 - (\tfrac{1}{32})k^4 - \cdots],$$

$$(3.7) \quad B_{02} = (Pgk^2/3\pi)[(\tfrac{3}{2}) + (\tfrac{1}{8})k^2 + (\tfrac{9}{256})k^4 + \cdots],$$

$$(3.8) \quad B_{40} = (Pg/15\pi)[-4 + 15k^2 - (\tfrac{225}{16})k^4 + \cdots],$$

$$(3.9) \quad B_{31} = (Pgk/3\pi)[-2 + (\tfrac{9}{4})k^2 - (\tfrac{15}{32})k^4 - \cdots],$$

$$(3.10) \quad B_{22} = (Pgk^2/\pi)[-(\tfrac{1}{2}) + (\tfrac{1}{8})k^2 + (\tfrac{5}{256})k^4 + \cdots],$$

$$(3.11) \quad B_{13} = (Pgk^3/3\pi)[-(\tfrac{1}{4}) - (\tfrac{3}{64})k^2 - (\tfrac{9}{512})k^4 - \cdots],$$

$$(3.12) \quad B_{04} = (Pgk^4/15\pi)[(\tfrac{5}{32}) + (\tfrac{9}{128})k^2 + (\tfrac{75}{2048})k^4 + \cdots],$$

where for the sake of simplicity the double signs on the subscripts have been omitted.

These series converge for all values of $k$, $0 < k = (Q/P) \leqq 1$, considered and thus in a sense provide a complete solution for the two-frequency modulation product problem for the full wave ideal rectifier. Similarly as in the special case of the zero biased ideal rectifier the coefficients (3.3) also may be expressed in finite form in terms of tabulated complete elliptic integrals for $k \neq 1$, or as rational functions of $\pi$ for $k = 1$, which in a sense thus provides another complete solution to the problem for the full wave ideal rectifier; see again the formulas (2.7) and (2.8) of Part II or the original results of Bennett [1; pp. 234–235] referred to earlier.

**4. The symmetric ideal limiter.** Consider next the symmetric ideal limiter having the output versus input characteristic

$$(4.1) \quad Y(X) = \begin{cases} -gX_0, & X \leqq -X_0, \\ gX, & -X_0 \leqq X \leqq X_0, \\ gX_0, & X \geqq X_0, \end{cases} \qquad 0 < X_0 < 2P, \qquad g > 0,$$

$X_0$ and $g$ being given constants as indicated. Again applying the theory of Part I, §3 we see that in any operating range $-a \leqq X \leqq a$, $a \geqq 2P$, the symmetric ideal limiter characteristic (4.1) may be expressed exactly in the form (3.1) of Part I with, in particular,

$$(4.2) \quad X_1 = -X_0, \qquad X_2 = X_0, \qquad g_1 = g, \qquad g_2 = -g, \qquad N = 2.$$

Hence, applying formulas (3.8) of Part I and simplifying with the aid of relations (4.8) of Part I much as before, we have the results

$$
\begin{aligned}
B_{\pm mn} &= 0, && (m+n = 0, 2, 4, \cdots),\\
B_{10} &= Pg[1 - 2A_{10}(h, k)],\\
B_{01} &= Pg[k - 2A_{01}(h, k)],\\
B_{\pm mn} &= -2PgA_{mn}(h, k), && (m+n = 3, 5, 7, \cdots),
\end{aligned} \tag{4.3}
$$

where $h = (X_0/P)$ and $k = (Q/P)$.

Similarly as before applying the double power series expansions of the functions $A_{mn}(h, k)$ and the recursion identities of Part I, §§4 and 5 we obtain for values of $h$ and $k$ in case (ii) of Part I, §4 the double power series expansions

$$
B_{10} = (Pgh/\pi)\left\{\begin{matrix} 4 - (3/5)h^2 - (1/10)h^4 - \cdots \\ - k^2 - (1/2)h^2k^2 - \cdots \\ - (3/16)k^4 - \cdots \\ \cdots \end{matrix}\right\}, \tag{4.4}
$$

$$
B_{01} = (Pghk/\pi)\left\{\begin{matrix} 2 + (1/3)h^2 + (3/20)h^4 + \cdots \\ + (1/4)k^2 + (3/8)h^2k^2 + \cdots \\ + (3/32)k^4 + \cdots \\ \cdots \end{matrix}\right\}, \tag{4.5}
$$

$$
B_{30} = (Pgh/\pi)\left\{\begin{matrix} - (1/8) + 2h^2 - (1/2)h^4 - \cdots \\ + 3k^2 - (5/2)h^2k^2 - \cdots \\ - (15/16)k^4 - \cdots \\ \cdots \end{matrix}\right\}, \tag{4.6}
$$

$$
B_{21} = (Pghk/\pi)\left\{\begin{matrix} - 2 + h^2 + (1/4)h^4 + \cdots \\ + (3/4)k^2 + (5/8)h^2k^2 + \cdots \\ + (5/32)k^4 + \cdots \\ \cdots \end{matrix}\right\}, \tag{4.7}
$$

$$
B_{12} = (Pghk^2/\pi)\left\{\begin{matrix} - (1/2) - (1/4)h^2 - (3/16)h^4 - \cdots \\ - (1/8)k^2 - (5/16)h^2k^2 - \cdots \\ - (15/256)k^4 - \cdots \\ \cdots \end{matrix}\right\}, \tag{4.8}
$$

$$
B_{03} = (Pghk^3/\pi)\left\{\begin{matrix} (1/12) + (1/8)h^2 + (5/32)h^4 + \cdots \\ + (3/64)k^2 + (25/128)h^2k^2 + \cdots \\ + (15/512)k^4 + \cdots \\ \cdots \end{matrix}\right\}, \tag{4.9}
$$

each of which thus converges for $h + k \leqq 1$, i.e., for $X_0 + Q \leqq P$ and where again the double signs on the subscripts have been omitted.

The reader will note readily from these expansions the general behavior of the corresponding coefficients $B_{\pm mn}$ for small values of $h$ and $k$ together with several specific properties of them; e. g., for $h + k \leqq 1$ the coefficients $B_{10}$, $B_{01}$, $B_{20}$, $B_{21}$, $B_{12}$ and $B_{03}$ are odd functions of $h$ and contain as a factor the power $k^n$ where $n$ is maximal and is the second subscript on the coefficient. The formulas (4.3) together with these expansions and the tables of Part II essentially provide a practical solution for the two-frequency modulation product problem for the symmetric ideal limiter.

**5. The unsymmetric ideal limiter.** We consider finally the unsymmetric ideal limiter having the output versus input characteristic

$$(5.1)\qquad Y(X) = \begin{cases} 0, & X \leqq 0, \\ gX, & 0 \leqq X \leqq X_0, \\ gX_0, & X \geqq X_0, \end{cases} \qquad 0 < X_0 < 2P, \qquad g > 0,$$

$X_0$ and $g$ again being given constants as indicated. Proceeding similarly as before we apply the theory of Part I, §3 with

$$(5.2)\qquad X_1 = 0, \qquad X_2 = X_0, \qquad g_1 = g, \qquad g_2 = -g, \qquad N = 2,$$

and obtain after simplification the results

$$(5.3)\qquad \begin{aligned} B_{\pm mn} &= Pg[A_{mn}(0, k) - A_{mn}(h, k)], \quad (m + n = 0, 2, 4, \cdots), \\ B_{10} &= (Pg/2)[1 - 2A_{10}(h, k)], \\ B_{01} &= (Pg/2)[k - 2A_{01}(h, k)], \\ B_{\pm mn} &= -PgA_{mn}(h, k), \qquad (m + n = 3, 5, 7, \cdots), \end{aligned}$$

where again $h = (X_0/P)$ and $k = (Q/P)$.

Applying the results of Part I, §§4 and 5 once more we note for values of $h$ and $k$ in case (ii) of Part I, §4 the double power series expansions

$$(5.4)\qquad \tfrac{1}{2}B_{00} = (Pgh/2) - (Pgh^2/\pi)\left\{\begin{array}{l} (\tfrac{1}{2}) + (\tfrac{1}{24})h^2 + (\tfrac{1}{80})h^4 + \cdots \\ + (\tfrac{1}{8})k^2 + (\tfrac{3}{32})h^2k^2 + \cdots \\ + (\tfrac{9}{128})k^4 + \cdots \\ \cdots \end{array}\right\},$$

$$(5.5)\qquad B_{10} = (Pgh/\pi)\left\{\begin{array}{l} 2 - (\tfrac{1}{3})h^2 - (\tfrac{1}{20})h^4 - \cdots \\ - (\tfrac{1}{2})k^2 - (\tfrac{1}{4})h^2k^2 - \cdots \\ - (\tfrac{3}{32})k^4 - \cdots \\ \cdots \end{array}\right\},$$

$$(5.6)\qquad B_{01} = (Pghk/\pi)\left\{\begin{array}{l} 1 + (\tfrac{1}{6})h^2 + (\tfrac{3}{40})h^4 + \cdots \\ + (\tfrac{1}{8})k^2 + (\tfrac{3}{16})h^2k^2 + \cdots \\ + (\tfrac{3}{64})k^4 + \cdots \\ \cdots \end{array}\right\},$$

$$(5.7)\qquad B_{20} = (Pgh^2/\pi)\left\{\begin{array}{l} 1 - (1/4)h^2 - (1/24)h^4 - \cdots \\ - (3/4)k^2 - (5/16)h^2k^2 - \cdots \\ - (15/64)k^4 - \cdots \\ \quad \cdots \end{array}\right\},$$

$$(5.8)\qquad B_{11} = (Pgh^2k/\pi)\left\{\begin{array}{l} (1/2) + (1/8)h^2 + (1/16)h^4 + \cdots \\ + (3/16)k^2 + (15/64)h^2k^2 + \cdots \\ + (15/128)k^4 + \cdots \\ \quad \cdots \end{array}\right\},$$

$$(5.9)\qquad B_{02} = (Pgh^2k^2/\pi)\left\{\begin{array}{l} - (1/8) - (3/32)h^2 - (5/64)h^4 - \cdots \\ - (3/32)k^2 - (25/128)h^2k^2 - \cdots \\ - (75/1024)k^4 - \cdots \\ \quad \cdots \end{array}\right\},$$

$$(5.10)\qquad B_{30} = (Pgh/\pi)\left\{\begin{array}{l} - (2/3) + h^2 - (1/4)h^4 - \cdots \\ + (3/2)k^2 - (5/4)h^2k^2 - \cdots \\ - (15/32)k^4 - \cdots \\ \quad \cdots \end{array}\right\},$$

$$(5.11)\qquad B_{21} = (Pghk/\pi)\left\{\begin{array}{l} - 1 + (1/2)h^2 + (1/8)h^4 + \cdots \\ + (3/8)k^2 + (5/16)h^2k^2 + \cdots \\ + (5/64)k^4 + \cdots \\ \quad \cdots \end{array}\right\},$$

$$(5.12)\qquad B_{12} = (Pghk^2/\pi)\left\{\begin{array}{l} - (1/4) - (1/8)h^2 - (3/32)h^4 - \cdots \\ - (1/16)k^2 - (5/32)h^2k^2 - \cdots \\ - (15/512)k^4 - \cdots \\ \quad \cdots \end{array}\right\},$$

$$(5.13)\qquad B_{03} = (Pghk^3/\pi)\left\{\begin{array}{l} (1/24) + (1/16)h^2 + (5/64)h^4 + \cdots \\ + (3/128)k^2 + (25/256)h^2k^2 + \cdots \\ + (15/1024)k^4 + \cdots \\ \quad \cdots \end{array}\right\},$$

each of which again converges for $h + k \leqq 1$, i.e., for $X_0 + Q \leqq P$.

The reader again will note from these expansions the general behavior of the corresponding coefficients $B_{\pm mn}$ for small values of $h$ and $k$ together with such elementary properties as, for instance, that for $h + k \leqq 1$ the odd order coefficients $B_{10}$, $B_{01}$, $B_{30}$, $B_{21}$, $B_{12}$ and $B_{03}$ are odd functions of $h$ while the even order coefficients $B_{20}$, $B_{11}$ and $B_{02}$ are even functions of $h$ and as before each of

these coefficients contains as a factor the power $k^n$ where $n$ is maximal and is the second subscript on the coefficient. As before the formulas (5.3) together with these expansions and the tables of Part II essentially provide a practical solution for the problem at hand. This completes our treatment of rectifiers and limiters.

**6. A remark on terminology.** From the general form of the expansion (1.2) of the present paper or (3.11) of Part I, the user might be inclined to refer to $\frac{1}{2}B_{00}$ as the "direct current" term, to $B_{10}$ as the "amplitude of the first harmonic of the first fundamental", to $B_{01}$ as the "amplitude of the first harmonic of the second fundamental" and so forth. However, precisely speaking as pointed out by H. Kaufman, such terminology would be correct if and only if the ratio $r = (q/p)$ of the two input frequencies $p$ and $q$ is an irrational number. For if $r = (q/p)$ is rational then there are always additional contributions to these and other terms which may or may not be appreciable depending on the value of $r$. Thus, to cite an example, if $q = 2p$ then the so called "direct current" term contains not only the term $\frac{1}{2}B_{00}$ but also the terms $B_{21} \cos(2\theta_p - \theta_q)$, $B_{42} \cos(4\theta_p - 2\theta_q)$, $B_{63} \cos(6\theta_p - 3\theta_q)$ and so forth. Hence, for precise application of the techniques of this paper to the two-frequency modulation product problem with commensurable input frequencies these facts may need to be given consideration if the results are to agree with experiment to a high degree of accuracy.

## REFERENCES

1. Bennett, W. R., "New Results in the Calculation of Modulation Products", *Bell System Technical Journal*, vol. 12 (1933), pp. 228–243.
2. Bennett, W. R., "The Biased Ideal Rectifier", *Bell System Technical Journal*, vol. 26 (1947), pp. 139–169.
3. Sternberg, R. L. and Kaufman, H., "A General Solution of the Two-Frequency Modulation Product Problem. I.", *this Journal*, vol. 32 (1953), pp. 233–242.
4. Sternberg, R. L., "A General Solution of the Two-Frequency Modulation Product Problem. II.", *this Journal*, vol. 33 (1954), pp. 68–79.

Laboratory for Electronics, Inc.
Boston, Massachusetts

(Received July 10, 1953)

# A THEORY FOR HYDROFOILS OF FINITE SPAN*

Y. T. Wu

**Introduction.** The purpose of the present investigation is to study the hydrodynamic properties of a hydrofoil of finite span moving with constant velocity through deep water at a fixed distance beneath the free surface. The span of the hydrofoil is parallel to the plane of the free surface. This problem has recently attracted a great deal of attention from both the theoretical and the practical points of view. However, the idea of using the hydrofoil as an aid for locomotion over a water surface is certainly not new.

As early as 1898 Forlanini in Italy tried to use hydrofoils for the purpose of supporting high-speed boats. That is, the hull, with hydrofoils fastened underneath, when moving at sufficiently high speeds, would be lifted above the water surface and the familiar ship bow-waves would be replaced by much weaker waves due to the submerged hydrofoil system. The resisting force of the water is then considerably reduced. Consequently, with the same propulsive power, a hydrofoil-equipped watercraft can travel much faster than conventional ones. Another practical application of hydrofoils was made in 1911 by Guidoni[2] who replaced the ordinary floats of a seaplane by hydrofoils. The use of hydrofoils as floats provides greatly improved aerodynamic characteristics of the seaplane during flight. However, the lack of knowledge of the fundamental properties and design data of hydrofoils has prevented the use of hydrofoils in modern applications. The recent revival of interest in the theoretical and experimental studies of this problem is aimed at removing these gaps in basic information.

It is known that, for a wing moving in an infinite fluid medium, one may neglect the influence of gravity and consider only the inertia and viscous effects; the Reynolds condition for dynamical similarity then holds. Under this condition, the nondimensional lift and drag coefficients depend only on the Reynolds number and the geometry of the body. However, the situation for a hydrofoil near the water surface is quite different. The hydrofoil differs from the airfoil not only because of the possible occurrence of cavitation but also through the strong effect of the free water surface. For the motion of a hydrofoil at shallow submergence, one must consider the gravity effect because the wave formations on the free water surface will influence decisively its hydrodynamic properties. It follows that, in this case, the nondimensional lift and drag coefficients will be functions not only of the Reynolds number and its geometry, but also of the Froude number and the cavitation number.

The program of this paper is as follows: After a brief survey of the available theoretical and experimental information on the characteristics of hydrofoils, the theory for a hydrofoil of finite span will be formulated. The liquid medium

* This paper is the somewhat condensed version of a Hydrodynamics Laboratory report of the same title.[1] The work was sponsored by the Office of Naval Research under the Contract Number N6onr-24426 (NR062-083).

is assumed to be incompressible and nonviscous and of infinite depth. The basic concept of the analysis is patterned after the famous Prandtl wing theory of modern aerodynamics in that the hydrofoil of large aspect ratio may be replaced by a lifting line. The lift distribution along the lifting line is the same as the lift distribution, integrated with respect to the chord of the hydrofoil, along the span direction. The induced velocity field of the lifting line is then calculated by giving proper consideration to lift distribution along the lifting line, free water surface pressure condition and wave formation. The "local velocity" so determined for flow around each local section perpendicular to the span of the hydrofoil can be considered as that of a two-dimensional flow around a hydrofoil without free water surface. The only additional feature of the flow in this sectional plane is the modification of the geometric angle of attack, as defined by the undisturbed flow, to the so-called effective angle of attack on account of the local induced velocity. Thus the local sectional characteristics to be used can be taken as those of a hydrofoil section in two-dimensional flow without free water surface but may involve cavitation. More precisely, the hydrofoil section at any location of the span has the same hydrodynamic characteristics as if it were a section of an infinite span hydrofoil in a fluid region of infinite extent at a geometric angle of attack equal to $\alpha_e$, together with proper modification of the free stream velocity. Such characteristics may be obtained by theory or by experiment and should be taken at the same Reynolds number and cavitation number. With this separation of the three-dimensional effects and the two-dimensional effects, the effects of Froude number are singled out. Thus a systematic and efficient analysis of the hydrofoil properties can be made.

The lifting line theory used in this problem is developed through various stages to include the formulation of the direct or indirect problems and also the problem of minimum drag. Finally, detailed calculations of the lift and drag coefficients of a specified hydrofoil are carried out for the case of elliptical distribution of circulation. The effects of the free water surface and the wave formation are examined in detail. Needless to say, the neglect of viscous effects will cause the omission of the frictional drag in the total drag calculation. However, this error can be estimated separately.

The author wishes to express his gratitude to Professor H. S. Tsien for suggesting this problem and for his many inspiring discussions; he also wishes to thank Professors M. S. Plesset, A. Erdelyi and Dr. J. D. Cole for many useful suggestions.

**Survey of Previous Results.** The investigation of the properties of hydrofoils has been so far mostly experimental. Although there exist some approximate theories, they do not appear to show the effects of wave formation on the hydrodynamic characteristics of the hydrofoil, especially when the submergence is shallow. In Guidoni's work,[2] it is claimed that some of the advantages of hydrofoils over conventional ships or floats are better lift-drag ratio and less sensitivity to a rough water surface so that their use may result in a substantial decrease in the structural weight of the hull. Some early preliminary work on the hydrofoil

problem was carried out by Keldysch and Lawrentjew,[3] Kotschin[4] and Wladimirow[5] in ZAHI. Their studies are, in a broad sense, only an extension and modification of Lamb's work on the submerged cylinder[6] and Havelock's work on the sphere beneath a water surface.[7, 8] In Keldysch's work, which is a special case of Kotschin's, the problem of the two-dimensional hydrofoil was solved by replacing the wing by a circular cylinder with circulation. Wladimirow solved the problem of hydrofoils of finite span by replacing the hydrofoil by a horseshoe vortex of constant circulation along the span and by assuming that the free water surface remains flat no matter how close the hydrofoil is to the surface. The calculation then reduces to a two-dimensional problem in the Trefftz's plane behind the wing. This approximation is even poorer than the assumption of infinite Froude number. In 1935, a test of a single hydrofoil (the NACA 0.0009 profile) was carried out by Wladimirow and Frolow in the towing tank of ZAHI.[8] Experimentally it was found that both lift and drag coefficients decrease with the depth of immersion, but the rate of decrease of the lift coefficient was faster than that of drag, especially for small depths. The comparison between theory and experiment on this point was not satisfactory. About the same time, the problem of cavitation on a submerged obstacle of hydrofoil section such as a ship propeller was approached by Ackeret[9], Walchner[10], Lerbs[11], Martyrer[12], Gutsche[13], and Smith[14]. Around 1937, further contributions to the hydrofoil problem were made by Weinig[15], Tietjens[16], and von Schertel[17]. Weinig gave a preliminary yet exhaustive discussion on hydrofoil and planing problems neglecting completely the effects of gravity and the free surface elevation; the wave resistance for both cases was estimated approximately afterwards. During the last World War, a series of tests on a number of hydrofoils of different foil sections was carried out at NACA by Land, Benson and Ward [18, 19, 20] in the NACA towing tank. Several interesting results were found. At depths greater than 4 chords, the influence of the free surface is negligibly small. In the range of depths between 4 chords and approximately $\frac{1}{2}$ chord, lift and drag coefficients decrease, and the cavitation speed increases, with decrease in depth until the hydrofoil approaches and breaks through the surface at which point a sudden decrease in lift occurs. However, the corresponding values of the lift-drag ratio increase to maximum values when the hydrofoils are near the surface then decrease rapidly with further decrease in depth to values for the planing surfaces. Obviously, these observations indicate that there exists an optimum value of the depth for which the hydrofoil operates most efficiently. The effect of an increase of speed is to reduce the minimum drag coefficient and to make it occur at a corresponding higher value of the lift coefficient until cavitation takes place. Any lower surface cavitation reduces the lift; while complete upper surface cavitation prevents further increase of lift. Under such condition, however, a maximum value of lift as high as 1 ton per square foot was recorded. A profile with sharp leading edge seems desirable for reducing cavitation. The hydrofoil section NACA 16-509 was observed to have favorable hydrodynamic properties. The stability problem of a system of hydrofoils was studied by Imlay[21]. An exhaustive experimental investigation of a single hydrofoil with large dihedral

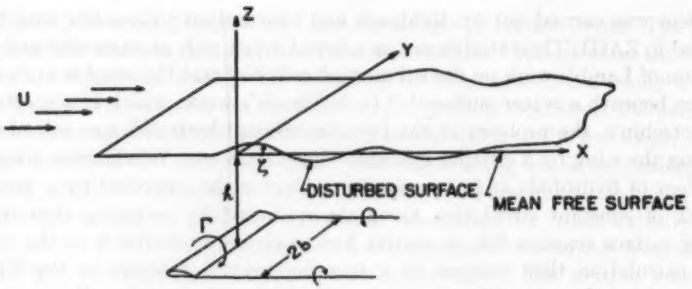



FIG. 1. Coordinate system for the motion of hydrofoil.

was made by Sottorf[22] in the Göttingen towing tank in which eighteen different profiles were studied; his results have been made available only recently. He found that a thin profile of almost circular segment form with pointed nose and an increased convex camber in the nose area is favorable for laminar flow, and consequently also reduces cavitation, gives higher lift and lift-drag ratio, and minimum spray formation at the junction of the tip of the hydrofoil and the free water surface. The main feature of the depth effects on hydrodynamic properties are in good agreement with NACA results. It was also found that a partial cavitation on the upper surface has a favorable effect provided that it covers less than a half chord because a thin cavitation bubble layer reinforces the flow curvature and thereby increases the lift. In addition, the bubble layer lessens the surface friction by acting as a cushion between the solid surface and the high velocity water stream. However, this favorable range of operation is not very stable.

**General Formulation of the Problem.** Consider a hydrofoil of span $2b$, with arbitrary profile and plan form, moving with constant forward velocity $U$ through deep water at fixed depth of immersion $h$ which is measured from its trailing edge to the mean free water surface. If we choose a coordinate system fixed with respect to the hydrofoil, then in this system the flow picture would appear to be stationary with a uniform free stream velocity $U$ approaching the hydrofoil. Let the $x$-axis be parallel to the direction of the free stream, the $z$-axis point upward, with $z = 0$ representing the undisturbed free surface and the trailing edge of the hydrofoil lying between $(0, -b, -h)$ and $(0, b, -h)$ as shown in Fig. 1. The elevation of the disturbed free surface caused by the hydrofoil is denoted by $\zeta(x, y)$ measured from $z = 0$. The liquid medium is assumed to be incompressible and nonviscous so that the condition of irrotationality and continuity implies that the perturbation velocity potential $\varphi$ satisfies the Laplace equation

$$\nabla^2\varphi = 0. \tag{1}$$

From its solution the total velocity field of the liquid flow can be obtained as

$$\mathbf{Q} = U\mathbf{i} + \mathbf{q}, \qquad \mathbf{q} = \operatorname{grad}\varphi \tag{2}$$

The pressure field may be determined from the Bernoulli integral

$$p/\rho + \tfrac{1}{2}\mathbf{Q}\cdot\mathbf{Q} + gz = \text{const.} \tag{3}$$

in which the effect of gravity has been taken into account.

Now let us assume that the resulting motion is such that almost everywhere $|\mathbf{q}|$ is much less than $U$ so that Eq. (3) may be linearized to yield:

$$p/\rho + U\varphi_x + gz = \text{const.} \tag{4}$$

The corresponding linearized boundary conditions on the disturbed, free surface now become

$$(\partial\varphi/\partial z)_{z=0} = U\partial\zeta/\partial x, \tag{5}$$

and

$$U(\partial\varphi/\partial x)_{z=0} = -g\zeta. \tag{6}$$

Eq. (5) represents the linearized kinematic condition on the free surface, while Eq. (6) is identical with the dynamic condition $p = p_0 =$ constant on the disturbed surface. It may be noted that gravity affects the kinematics through the boundary conditions (5) and (6) and thereafter influences the dynamics through Eq. (4), even though it does not appear explicitly in the kinematical Eq. (1). The boundary condition on the hydrofoil surface is

$$\partial\varphi/\partial n = -U\cos(n, x) \tag{7}$$

where $\partial\varphi/\partial n$ is the normal component of the velocity at the surface. It may also be mentioned here that the real physical situation requires that the disturbance upstream should diminish at a rate which is made to agree with observation.

In order to simplify the problem, it will be assumed that the aspect ratio

$$AR = (2b)^2/S \tag{8}$$

of the hydrofoil with plan area $S$ and span $2b$ is so large that the whole hydrofoil may be replaced by a lifting line. Then the subsequent analysis can be made similar to the Prandtl wing theory of modern aerodynamics by introducing appropriate modifications of Prandtl's original concept. The fundamental concept is that the circulation distribution along the lifting line which replaces the hydrofoil is the same as the distribution of bound vorticity, integrated with respect to the chord of the hydrofoil, along the span direction. We then calculate the induced velocity field of this circulation distribution on the lifting line. The characteristic length of this induced velocity field is the span 2b or the immersion depth $h$, whichever is the smaller. Therefore if $h$ is much larger than the representative chord of the hydrofoil, then the characteristic length of the induced velocity field is very much larger than the characteristic length of the local velocity field which is associated with the bound vorticity and the circulation. Then as far as the lift production is concerned, each section of the hydrofoil is effectively surrounded by a stream of infinite extent moving with an effective velocity which is not the free stream velocity $U$ but the sum of $U$ and the induced velocity. We then replace the boundary condition of Eq. (7) by the

statement that the circulation of the lifting line at a certain spanwise station is the same as that corresponding to a two-dimensional flow without free surface around the same local section with an effective free stream velocity equal to the sum of $U$ and the induced velocity. Although our argument for the lifting line concept is based upon large aspect ratio $\mathcal{R}$, experience with airfoils indicates that the theory can be expected to be sufficiently accurate even for $\mathcal{R}$ as low as four.

We shall assume that along the lifting line the distribution of the circulation strength $\Gamma(y)$ is known for given angle of attack and immersion depth. It should be pointed out here that for shallow submergence the effects of the free surface modify considerably the approaching flow velocities and thereby make the pressure and lift distribution different from its corresponding aerodynamic value. Therefore, in our hydrofoil problem the lift distribution is no longer simply proportional to $\Gamma(y)$ as it is in aerodynamic wing theory but is rather a complicate function of $\Gamma(y)$. When $\Gamma(y)$ is given, the lift distribution can be determined as will be shown later. However, the calculation of $\Gamma(y)$ for known lift distribution is more difficult. Strictly speaking, $\Gamma(y)$ depends not only on the flow velocity, angle of attack and the geometry of the hydrofoil, but also on the depth of submergence $h$. When $h$ tends to infinity so that this hydrofoil is in a flow of infinite extent, the free surface and wave effects should vanish and $\Gamma(y)$ then tends to its conventional aerodynamic value $\Gamma_\infty(y)$. Since the circulation for equal airfoil and equal effective angle of attack is proportional to the approaching relative flow velocity, we have

$$\Gamma(y) = [1 + U^{-1}\partial\varphi(0, y, -h)/\partial x]\Gamma_\infty(y) \tag{9}$$

If $\Gamma_\infty(y)$ instead of $\Gamma(y)$ at depth $h$, were given, Eq. (9) would lead to an integral equation for $\Gamma(y)$ as will be shown later. A similar situation results if one intends to find $\Gamma(y)$ at $h$ for given geometry of the hydrofoil. Therefore we shall first consider the simple case in which $\Gamma(y)$ is assumed to be given.

In this case, the contribution of the lifting line to the perturbation potential can be shown to have the following integral representation (cf. Ref. 23, also cf. Ref. 1, Appendix I):

$$\varphi_1 = \frac{\text{sign } (z+h)}{2\pi^2}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^\infty\frac{\sin\lambda x}{\lambda}\,d\lambda\int_0^\infty\cos\mu(y-\eta)e^{-\sqrt{\lambda^2+\mu^2}\,|z+h|}\,d\mu + \frac{\text{sign } (z+h)}{4\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^\infty\cos\mu(y-\eta)e^{-\mu|z+h|}\,d\mu. \tag{10}$$

It may be noted that the above expression satisfies Eq. (1) and has a singularity along the lifting line, a jump of $\Gamma(y)$ across the vortex sheet on the downstream side of the lifting line and is regular elsewhere inside the water. Needless to say, $\varphi_1$ has no meaning above the free surface ($z > \zeta$). It may also be remarked here that some other useful representations for $\varphi_1$ can be written as follows (cf. Ref. 1, Appendix I):

$$\varphi_1 = \frac{(z+h)}{4\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^\infty d\xi\int_0^\infty e^{-k|z+h|}J_0(k\sqrt{(x-\xi)^2+(y-\eta)^2})k\,dk, \tag{10a}$$

or,

$$\varphi_1 = \frac{(z+h)}{4\pi}\int_{-\infty}^{\infty}\frac{\Gamma(\eta)}{(y-\eta)^2+(z+h)^2}\cdot\left[1+\frac{x}{\sqrt{x^2+(y-\eta)^2+(z+h)^2}}\right]d\eta, \tag{10b}$$

where $J_0$ denotes the Bessel function of the first kind.

It is convenient to decompose $\varphi$ into three parts

$$\varphi = \varphi_1 + \varphi_2 + \varphi_3 \tag{11}$$

in such a way that the boundary condition of Eq. (6) is decomposed as follows:

$$\partial\varphi_1/\partial x + \partial\varphi_2/\partial x = 0 \qquad \text{at } z = 0, \tag{12}$$

and

$$U(\partial\varphi_3/\partial x)_{z=0} = -g\zeta. \tag{13}$$

Physically this decomposition means that $\varphi_3$ represents the potential due to wave formation so that $\varphi_3$ is the only part of $\varphi$ which involves the effect of gravity, and it vanishes when the gravitational effect is neglected. In the latter case the boundary condition Eq. (6) reduces to Eq. (12). It does not follow, however, that the disturbance $\zeta$ should also vanish because $\zeta$ is also influenced by $\varphi_1$ and $\varphi_2$ through Eq. (5). The remaining part of $\varphi$, namely, $\varphi_1 + \varphi_2$, is equivalent to the potential of a biplane system in infinite flow with the upper lifting line of an equal circulation $\Gamma(y)$, rotating in the same sense, distributed along the image points ($z = +h$) of the real wing. In other words, $\varphi_2$ represents the correction to $\varphi$ due to the effect of the mean free surface $z = 0$. According to this reflection, $\varphi_2$ may be written down directly from Eq. (10) by replacing $h$ by $(-h)$. In the range of present interest, we have

$$\varphi_2 = -\frac{1}{2\pi^2}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}\frac{\sin\lambda x}{\lambda}\,d\lambda\int_0^{\infty}\cos\mu(y-\eta)e^{-\sqrt{\lambda^2+\mu^2}(z-h)}\,d\mu$$
$$-\frac{1}{4\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}\cos\mu(y-\eta)e^{\mu(z-h)}\,d\mu, \qquad \text{for } z \leqq 0, \tag{14}$$

which may also be converted into expressions similar to Eqs. (10a) and (10b). It may be noted that this expression is regular everywhere inside the water.

Now it seems plausible to assume that $\varphi_3$ and $\zeta$ may be represented by the following expressions:

$$\varphi_3 = \int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}\alpha(\lambda,\mu)\frac{\sin\lambda x}{\lambda}\,d\lambda\int_0^{\infty}\cos\mu(y-\eta)e^{\sqrt{\lambda^2+\mu^2}(z-h)}\,d\mu$$
$$+\frac{1}{2\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}\cos\mu(y-\eta)e^{\mu(z-h)}\,d\mu, \qquad \text{for } z \leqq 0;$$

$$\zeta = \int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}\beta(\lambda,\mu)\cos\lambda x\,d\lambda\int_0^{\infty}\cos\mu(y-\eta)e^{-\sqrt{\lambda^2+\mu^2}\lambda}\,d\mu;$$

where $\alpha(\lambda, \mu)$ and $\beta(\lambda, \mu)$ are two functions of $\mu$ and $\lambda$ to be determined by using boundary conditions (5) and (13). Substituting these values into Eqs. (5) and (13), we obtain:

$$\alpha(\lambda, \mu) = -\frac{\kappa}{\pi^2[\lambda^2(\lambda^2 + \mu^2)^{-\frac{1}{2}} - \kappa]}, \qquad \kappa = \frac{g}{U^2},$$

and

$$\beta(\lambda, \mu) = -(U/g)\alpha(\lambda, \mu).$$

These integral representations of $\varphi_3$ and $\zeta$ can be converted into forms convenient for calculation by introducing the new variables $k$ and $\theta$ such that:

$$\lambda = k \cos\theta, \qquad \mu = k \sin\theta.$$

Then

$$\varphi_3 = -\frac{\kappa}{\pi^2}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi}d\xi\int_0^{\pi/2}\sec^2\theta\,d\theta\int_0^{\infty}e^{k(z-\lambda)} \cdot\frac{\cos(k\xi\cos\theta)\cos(k(y-\eta)\sin\theta)}{k-\kappa\sec^2\theta}\,k\,dk \tag{15}$$

$$+\frac{1}{2\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}e^{\mu(z-\lambda)}\cos\mu(y-\eta)\,d\mu, \qquad \text{for } z \leqq 0;$$

$$\zeta = \frac{1}{\pi^2 U}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi/2}\sec^2\theta\,d\theta\int_0^{\infty}e^{-\lambda k} \cdot\frac{\cos(kx\cos\theta)\cos(k(y-\eta)\sin\theta)}{k-\kappa\sec^2\theta}\,k\,dk. \tag{16}$$

It can be seen that this value of $\zeta$ given by Eq. (16) is even in $x$ so that the elevation upstream is merely the reflection of that downstream. Therefore it seems necessary to investigate the behavior of $\zeta$ for large negative values of $x$ in order to make a comparison with observation. The integral representing $\zeta$ as given by Eq. (16) is indeterminate, but it is familiar in plane wave problems that its principal value can be obtained by considering $k$ to be complex and evaluating it by contour integration. The path on the positive real axis should be indented around the point $k = \kappa\sec^2\theta$ which is a simple pole of the integrand. After deforming the contour to the imaginary axis, we obtain:

$$\text{Principal value of } \int_0^{\infty}e^{-\lambda k}\frac{\cos(kx\cos\theta)\cos(k(y-\eta)\sin\theta)}{k-\kappa\sec^2\theta}\,k\,dk$$

$$= -(\operatorname{sign} x)\pi\kappa\sec^2\theta\,e^{-k\kappa\sec^2\theta}\sin(\kappa x\sec\theta)\cos(\kappa(y-\eta)\sec^2\theta\sin\theta) \tag{17}$$

$$+\int_0^{\infty}e^{-k|x|\cos\theta}\cosh(k(y-\eta)\sin\theta)\frac{\kappa\sec^2\theta\cos(kh)+k\sin(kh)}{k^2+\kappa^2\sec^4\theta}\,k\,dk.$$

The integration of the last integral in Eq. (17) with respect to $\theta$ (cf. Eq. (16)) should only extend over the range for which $|x|\cos\theta \pm (y-\eta)\sin\theta > 0$ for

an assigned point $(x, y)$. It is evident from Eq. (17) that although the integral vanishes as $|x| \to \infty$, the first term remains even for $x \to -\infty$. Then there are surface waves even far ahead of the hydrofoil. This is in contradiction with experience. The paradox can be resolved by observing that there are solutions of the basic flow equation, the free wave solutions, which satisfy the surface pressure condition of Eq. (6). We can then add these solutions to our velocity potential without rendering the solution so far obtained invalid for our problem. The appropriate free wave solution to be added is determined by the condition that the free surface elevation $\zeta$ must vanish at large distances ahead of the hydrofoil. Therefore, we have to superimpose on $\zeta$ another system of waves

$$\zeta' = -\frac{4\kappa}{\pi U}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi/2}\sec^4\theta\,e^{-h\kappa\sec^2\theta}\sin(\kappa x\sec\theta)\cdot\cos(\kappa(y-\eta)\sec^2\theta\sin\theta)\,d\theta \tag{18}$$

which is odd in $x$. In order to remain consistent with all the boundary conditions, we must add to the velocity potential another term, say, $\varphi_4$, given by

$$\varphi_4 = -\frac{\kappa}{\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi/2}\sec^3\theta\,e^{\kappa(z-h)\sec^2\theta}\cdot\cos(\kappa x\sec\theta)\cos(\kappa(y-\eta)\sec^2\theta\sin\theta)\,d\theta, \qquad (\text{for } z \leqq 0) \tag{19}$$

which obviously satisfies Eq. (1). This part of the potential may be thought of as a correction term. The same result can be obtained by introducing a fictitious viscous term to the Bernoulli equation (cf. Ref. 1 Appendix II, (Eq. (18)), a device first discovered by Lord Rayleigh. The physical significance of this viscous term is, however, not easy to understand. Our argument, even if somewhat lengthy, has its merit.

From the resulting expression of the surface elevation, $\zeta + \zeta'$, it can be seen that the term with the factor $e^{-h|x|\cos^2\theta}$ in Eq. (17) diminishes exponentially with increasing $|x|$ both in upstream and downstream direction. Hence for large positive values of $x$, the surface elevation can be approximated by

$$\zeta \sim -\frac{2\kappa}{\pi U}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi/2}\sec^4\theta\,e^{-h\kappa\sec^2\theta}\cdot\sin(\kappa x\sec\theta)\cos(\kappa(y-\eta)\sec^2\theta\sin\theta)\,d\theta, \qquad (\text{for } x > 0, \text{ large}) \tag{20}$$

This relation shows that $\zeta$ is analyzed into components of simple waves, where each train of waves (corresponding to one value of $\theta$) propagates on the downstream side of the wave front $x\cos\theta + (y-\eta)\sin\theta =$ constant with wave length

$$l = \frac{2\pi}{\kappa\sec^2\theta} = \frac{2\pi U^2}{g}\cos^2\theta. \tag{21}$$

This configuration of surface waves resembles that of typical ship waves (cf. Ref. 6, pp. 433–437). To study the behavior of $\zeta$ behind the hydrofoil as $x \to +\infty$, $y$ finite, we can approximate the integral in Eq. (20) by applying the

method of stationary phase (e.g. cf. Ref. 24, p. 505). Only the result will be given here:

$$\zeta \sim -\frac{1}{U^2}\sqrt{\frac{2g}{\pi x}}\sin\left(\kappa x + \frac{\pi}{4}\right)\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta + O\left(\frac{1}{x}\right), \quad \text{as } x \to +\infty,\ y \text{ finite.} \tag{22}$$

Therefore $\zeta$ tends to zero like $O(x^{-\frac{1}{2}})$ far downstream; and at fixed $x$ the effect of free stream velocity is such that $\zeta$ is proportional to $U^{-1}$, noting that in Eq. (22) $\Gamma(y)$ is proportional to $U$. Eq. (22) shows that as the distance from the hydrofoil is increased only the total circulation on the lifting line matters, the detail of the distribution of circulation is inconsequential. This is, of course, what we would expect from general principles.

It may also be seen from Eqs. (16) and (18) that there is a local disturbance immediately above the obstacle. As we do not intend to examine the surface wave pattern in detail, the above discussion suffices to describe the general behavior of the surface elevation.

Thus far we have found the complete perturbation potential, namely,

$$\varphi = \varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 \tag{23}$$

where $\varphi_1$, $\varphi_2$, $\varphi_3$ and $\varphi_4$ are given by Eqs. (10), (14), (15) and (19) respectively.

**Calculation of the Drag from Induced Velocities.** The total drag, neglecting the part caused by frictional effects, experienced by the hydrofoil may be calculated by using the Kutta-Joukowsky law:

$$D = -\rho\int_{-\infty}^{\infty}\Gamma(y)\,\frac{\partial\varphi(0, y, -h)}{\partial z}\,dy \tag{24}$$

where $-(\partial/\partial z)\varphi(0, y, -h)$ represents the total induced downwash at the trailing edge of the hydrofoil. If we split $D$ into components, each of which corresponds to the respective component of $\varphi$, we have

$$D = D_1 + D_2 + D_3 + D_4$$

where

$$D_n = -\rho\int_{-\infty}^{\infty}\Gamma(y)(\partial/\partial z)\varphi_n(0, y, -h)\,dy, \qquad (n = 1, 2, 3, 4) \tag{25}$$

The calculation of $D_1$, the induced drag due to the trailing vortices of the hydrofoil itself, is familiar in three-dimensional aerodynamic wing theory and is given by (cf. Ref. 23):

$$D_1 = \frac{\pi\rho}{4}\int_0^{\infty}[f(\mu)^2 + g(\mu)^2]\,\mu\,d\mu \tag{26}$$

where

$$f(\mu) = \frac{1}{\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\cos\mu\eta\,d\eta, \qquad g(\mu) = \frac{1}{\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\sin\mu\eta\,d\eta, \tag{27}$$

$f(\mu)$ and $g(\mu)$ are thus merely the Fourier-coefficients of the circulation distribution $\Gamma(y)$.

For the rest of $D$, we first calculate $\partial\varphi_a/\partial z$ from Eqs. (14), (15) and (19) and obtain:

$$\left(\frac{\partial\varphi_2}{\partial z}+\frac{\partial\varphi_3}{\partial z}\right)_{z=0,z=-h}=\frac{1}{4\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}e^{-2h\mu}\cos\left(\mu(y-\eta)\right)\mu\,d\mu,$$

$$\left(\frac{\partial\varphi_2}{\partial z}\right)_{z=0,z=-h}=-\frac{\kappa^2}{\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi/2}\sec^5\theta\,e^{-2h\kappa\sec^2\theta}\cos\left(\kappa(y-\eta)\sec^2\theta\sin\theta\right)d\theta.$$

Substitution of these values into Eq. (25) leads to:

$$D_2+D_3=-\frac{\pi\rho}{4}\int_0^{\infty}e^{-2h\mu}[f(\mu)^2+g(\mu)^2]\mu\,d\mu, \tag{28}$$

and

$$D_4=\pi\rho\kappa^2\int_0^{\pi/2}e^{-2h\kappa\sec^2\theta}[f(\kappa\sec^2\theta\sin\theta)^2+g(\kappa\sec^2\theta\sin\theta)^2]\sec^5\theta\,d\theta \tag{29}$$

where $f$ and $g$ are defined by Eq. (27). The combination of $D_2$ and $D_3$ represents the total contribution of the mean free surface effect, which favorably decreases the total drag especially when $h$ is small. As $h$ tends to zero, $D_2+D_3$ cancels $D_1$. The part $D_4$ represents the wave drag due to the downstream wave formation which results from the gravity effect. It may be remarked here that the wave drag can also be obtained by the method of traveling surface pressure (cf. Ref. 1 Appendix III).

From the above result we see that the drag is expressed in terms of the Fourier coefficients, ($f(\mu)$ and $g(\mu)$), of the circulation distribution $\Gamma(y)$. Therefore for a given $\Gamma(y)$, the total drag is completely determined. In most cases, the wing form is symmetric with respect to the central plane $y=0$, that is,

$$\Gamma(y)=\Gamma(-y), \tag{30}$$

then it follows from Eq. (27) that $g$ vanishes identically under this condition.

By substituting the relations of Eq. (27) into Eqs. (26), (28) and (29), the drags can be expressed directly in terms of the circulation distribution. Thus

$$D_1=-\frac{\rho}{4\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_{-\infty}^{\infty}\frac{d\Gamma(\eta')/d\eta'}{\eta'-\eta}\,d\eta', \tag{26a}$$

$$D_2+D_3=-\frac{\rho}{4\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Gamma(\eta)\Gamma(\eta')\frac{(2h)^2-(\eta-\eta')^2}{[(2h)^2+(\eta-\eta')^2]^2}\,d\eta\,d\eta' \tag{28a}$$

and

$$D_4=\rho\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Gamma(\eta)\Gamma(\eta')G(\eta-\eta')\,d\eta\,d\eta' \tag{29a}$$

where

$$G(\eta - \eta') = \frac{\kappa^2}{\pi} \int_0^{\pi/2} e^{-2h\kappa \sec^2\theta} \cos\left[(\eta - \eta')\kappa \sec^2\theta \sin\theta\right] \sec^5\theta \, d\theta \tag{29b}$$

Eq. (26a) is the familiar drag formula for the Prandtl lifting line theory. Eq. (29b) shows that $G(\eta - \eta')$ is symmetrical with respect to $\eta$ and $\eta'$.

It is seen that the drag components $D_1$, $D_2$ and $D_3$ are independent of the gravitational effect. They are thus properly called the components of the *induced drag*. All the drag caused by the presence of surface waves is represented by $D_4$. $D_4$ is thus the *wave drag*. Our calculation does not include, however, the skin friction drag produced by the viscous shearing stress on the surface of the hydrofoil. This skin friction drag should be added to the drag calculated in this section to obtain the total drag of the hydrofoil.

**Calculation of the Total Lift.** It is well known that the lift exerted on a wing flying in an infinite fluid is given by

$$L_0 = \rho U \int_{-\infty}^{\infty} \Gamma(y) \, dy \tag{31}$$

In the hydrofoil problem, however, the approaching stream velocity is influenced by effects of the free surface and wave formations. Hence, the total lift on the hydrofoil will differ from its aerodynamic value $L_0$ by an amount $\Delta L$ such that the total lift

$$L = L_0 + \Delta L \tag{32}$$

where

$$\Delta L = \rho \int_{-\infty}^{\infty} \Gamma(y) \frac{\partial \varphi(0, y, -h)}{\partial x} \, dy \tag{33}$$

From the expression of $\varphi$ we find that $\varphi_1$ and $\varphi_4$ have no contribution to the value of $(\partial\varphi/\partial x)$ at the hydrofoil. The final result is

$$\begin{aligned} \frac{\partial \varphi(0, y, -h)}{\partial x} &= \left(\frac{\partial \varphi_2}{\partial x} + \frac{\partial \varphi_3}{\partial x}\right) \quad \text{at} \quad x = 0, z = -h \\ &= -\frac{1}{2\pi^2} \int_{-\infty}^{\infty} \Gamma(\eta) \, d\eta \int_0^{\pi/2} d\theta \int_0^{\infty} e^{-2hk} \cos\left(k(y - \eta) \sin\theta\right) \\ &\qquad \cdot \frac{k + \kappa \sec^2\theta}{k - \kappa \sec^2\theta} \, k \, dk \end{aligned} \tag{34}$$

Substitution of this value into Eq. (33) yields:

$$\Delta L = -\frac{\rho}{2} \int_0^{\infty} e^{-2h\mu} \mu \, d\mu \int_0^{\pi/2} \left[f(\mu \sin\theta)^2 + g(\mu \sin\theta)^2\right] \frac{\mu + \kappa \sec^2\theta}{\mu - \kappa \sec^2\theta} \, d\theta \tag{35}$$

where $f$ and $g$ are again given by Eq. (27) and $g$ vanishes for symmetrical wings. Equation (35) represents the contribution of free surface effects, which tend to lessen the total lift.

By substituting the relations of Eq. (34) into Eq. (33), a direct expression of $\Delta L$ in terms of the circulation distribution $\Gamma(y)$ is obtained:

$$\Delta L = -\frac{\rho}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Gamma(\eta)\Gamma(\eta')F(\eta-\eta')\,d\eta\,d\eta' \tag{35a}$$

where

$$F(\eta-\eta') = \frac{1}{\pi^2}\int_0^{\infty} e^{-2h\mu}\,d\mu\int_0^{\pi/2}\cos\left[(\eta-\eta')\mu\sin\theta\right]\frac{\mu+\kappa\sec^2\theta}{\mu-\kappa\sec^2\theta}\,d\theta \tag{35b}$$

Therefore the magnitude of $\Delta L$ is in general equal to the magnitude of drag, i.e., of one order smaller than the lift $L_0$.

In connection with the present discussion, we examine the effect of immersion depth on the value of $\Gamma(y)$. Substituting Eq. (34) into Eq. (9), we obtain the following relation for a symmetrical hydrofoil,

$$\Gamma(y) = \Gamma_\infty(y)\left\{1-\frac{1}{2\pi U}\int_0^{\infty} e^{-2h\mu}\mu\,d\mu\int_0^{\pi/2} f(\mu\sin\theta)\cos(\mu y\sin\theta)\cdot\frac{\mu+\kappa\sec^2\theta}{\mu-\kappa\sec^2\theta}\,d\theta\right\} \tag{36}$$

where $\Gamma_\infty(y)$ is the value of $\Gamma(y)$ for the same wing and same effective angle of attack as $h\to\infty$. If $\Gamma(y)$ at depth $h$ is given, then the aerodynamic value of the circulation of the same wing, can be calculated from Eq. (36). On the other hand, if only $\Gamma_\infty(y)$ is known, then Eq. (36) provides an integral equation for $\Gamma(y)$ since $f$ also depends on $\Gamma(y)$. However, for moderate values of $h$, $\Gamma(y)$ and $\Gamma_\infty(y)$ are approximately equal as will be shown later.

**Geometry of the Hydrofoil; Effective Angle of Attack.** Take a strip of the hydrofoil of width $(dy)$ in the spanwise direction and with chord $c(y)$ located at $y$. According to the basic concepts of the lifting line theory explained in the section "General Formulation of the Problem", the circulation $\Gamma(y)$ around the lifting line at $y$ is the same as the circulation around the corresponding hydrofoil section. It then follows that the lift on this lifting line segment is the same as the lift on this hydrofoil section, integrated with respect to the chord of the hydrofoil. The local flow around this hydrofoil section in a plane $y$ = constant may be then considered as a two-dimensional flow around the same hydrofoil section, with free stream velocity $(U+u(y), w(y))$ and without free water surface. Therefore the local lift coefficient $C_l(y)$ and drag coefficient $C_d(y)$ per unit chord length can be defined by

$$\tfrac{1}{2}\rho[U+u(y)]^2c(y)C_l(y)\,dy = \rho[U+u(y)]\Gamma(y)\,dy,$$

$$\tfrac{1}{2}\rho[U+u(y)]^2c(y)C_d(y)\,dy = \rho w(y)\Gamma(y)\,dy$$

where $u(y)$ and $w(y)$ denote the induced velocities $(\partial\varphi/\partial x)$ and $(-\partial\varphi/\partial z)$ respectively at the hydrofoil. The above relations then give

$$\Gamma(y) = \tfrac{1}{2}[U+u(y)]c(y)C_l(y) \tag{37}$$

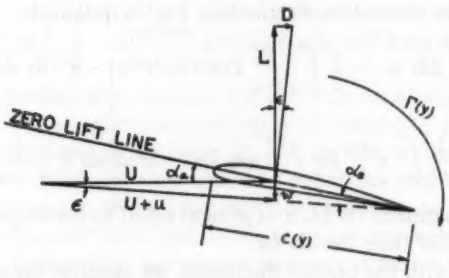



FIG. 2. Downwash and effective angle of attack.

and

$$\epsilon(y) = C_d/C_l = w(y)/[U + u(y)], \tag{38}$$

where $\epsilon(y)$ is defined by the above equation as the downwash angle. Eq. (38) indicates an important feature of the local flow that the actual absolute angle of attack $\alpha_a$, measured from the free stream to the zero-lift direction, is modified to an effective angle of attack $\alpha_e$, measured from the approaching stream to the zero-lift direction on account of the local perturbed velocity as shown in Fig. 2. The relation between $\alpha_a$ and $\alpha_e$ is given by

$$\alpha_e = \alpha_a - \epsilon(y) \tag{39}$$

In this way we can separate the three-dimensional effects and two-dimensional effects by taking the local sectional characteristics of the hydrofoil at depth $h$ and at absolute incidence $\alpha_a$ to be the same characteristics of the section of a two-dimensional hydrofoil submerged at infinite depth and sustained at an absolute incidence equal to $\alpha_e = \alpha_a - \epsilon(y)$. Referring to these two incidences, $C_l(y)$ is approximately proportional to $\alpha_a$ or $\alpha_e$ for most of the usual profiles at small incidences (Ref. 18 and 22). Moreover, when $\alpha_a$ vanishes so do $\alpha_e$ and $\epsilon(y)$. Hence we may put

$$C_l(y) = a_a \alpha_a = a_e \alpha_e . \tag{40}$$

In general, all $a_a$, $\alpha_a$, $a_e$ and $\alpha_e$ may be functions of $y$. In particular, $\alpha_a$ is constant for wings with no geometric twist, and in addition with the same profile along the span, $a_e$ is constant. Combining Eqs. (37)–(40), we have

$$\Gamma(y) = \frac{Uc(y)a_e}{2}\left[\alpha_a\left(1 + \frac{u(y)}{U}\right) - \frac{w(y)}{U}\right]. \tag{41}$$

For given $\Gamma(y)$, $u(y)$ and $w(y)$ are determined, then this relation gives the value of the chord length $c(y)$ except for a proportionality constant. In particular,

$$\Gamma(0) = \frac{Uc_0 a_e}{2}\left[\alpha_a\left(1 + \frac{u(0)}{U}\right) - \frac{w(0)}{U}\right] \tag{41a}$$

which gives the relation between $\Gamma(0)$ and $c_0$, the chord at the central section.

**Formulation of Problem with Specified Geometry.** The problem is for given geometry of the hydrofoil, that is, given $b$, $c(y)$, $\alpha_a(y)$, $a_o$ and $h$, to find $\Gamma(y)$, $C_L$ and $C_D$. The relation between the given quantities and the unknown $\Gamma(y)$ was approximated in the previous section to be

$$\frac{\Gamma(y)}{\frac{1}{2}Uc(y)a_o} = \alpha_a\left(1 + \frac{u(y)}{U}\right) - \frac{w(y)}{U}, \tag{42}$$

where from Eq. (34) and the section "Calculation of the Drag from Induced velocities",

$$u(y) = -\frac{1}{2\pi^2}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi/2}d\theta\int_0^{\infty}e^{-2h\mu}\cos\left[\mu(y-\eta)\sin\theta\right] \cdot \frac{\mu + \kappa\sec^2\theta}{\mu - \kappa\sec^2\theta}\mu\,d\mu, \tag{43}$$

$$w(y) = \frac{1}{4\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\infty}(1 - e^{-2h\mu})\cos\left[\mu(y-\eta)\right]\mu\,d\mu + \frac{\kappa^2}{\pi}\int_{-\infty}^{\infty}\Gamma(\eta)\,d\eta\int_0^{\pi/2}\sec^5\theta e^{-2h\kappa\sec^2\theta}\cos\left[\kappa(y-\eta)\sec^2\theta\sin\theta\right]d\theta \tag{44}$$

One way to solve this integral equation is by using a method similar to that of I. Lotz[28] in aerodynamic wing theory. This method consists in assuming for the circulation $\Gamma(y)$ at distance $y$ from the plane of symmetry the formula

$$\Gamma(y) = 4bU\sum_{n=0}^{\infty}A_{2n+1}\sin(2n+1)\phi, \qquad y = b\cos\phi; \tag{45}$$

$$= 0, \qquad |y| \geqq b$$

so that $\Gamma(y) = \Gamma(-y)$ and

$$\Gamma(0) = 4bU\sum_{n=0}^{\infty}(-)^nA_{2n+1}. \tag{45a}$$

The coefficients $A_{2n+1}$ are to be determined by using the condition (42). The first term of the Fourier series in Eq. (45), namely,

$$\Gamma_1(y) = 4bUA_1\sin\phi = 4bUA_1\sqrt{1 - \frac{y^2}{b^2}}, \tag{45b}$$

represents an elliptical distribution of the circulation. The rest of the terms may be regarded as a measure of the deviation between the actual distribution and the elliptical.

Substituting Eq. (45) in (43) and integrating by parts with respect to $\eta$, we obtain (cf. Ref. 1, Appendix IV, A)

$$u(y) = -U\sum_{n=0}^{\infty}(-)^n(2n+1)A_{2n+1}\mathfrak{U}_{2n+1}(\phi) \tag{46}$$

where

$$\mathfrak{U}_{2n+1}(\phi) = \frac{2}{\pi}\int_0^{\pi/2}\frac{d\theta}{\sin\theta}\int_0^{\infty}e^{-2\lambda t}\frac{t + b\kappa\sec^2\theta}{t - b\kappa\sec^2\theta}J_{2n+1}(t\sin\theta)\cos(t\cos\phi\sin\theta)\,dt, \tag{46a}$$

$$\lambda = \frac{h}{b} = \text{depth} - \tfrac{1}{2}\text{ span ratio}. \tag{46b}$$

In a similar manner, substitution of (45) in (44) gives (cf. Ref. 1, Appendix IV, B)

$$w(y) = \frac{U}{\sin\phi} \sum_{n=0}^{\infty} (2n+1) A_{2n+1} \mathfrak{W}_{2n+1}(\phi) \tag{47}$$

where

$$\mathfrak{W}_{2n+1}(\phi) = \sin(2n+1)\phi + (-)^{n+1}(\sin\phi) Rl \cdot \frac{[\sqrt{1+(2\lambda - i\cos\phi)^2} - (2\mu - i\cos\phi)]^{2n+1}}{\sqrt{1+(2\lambda - i\cos\phi)^2}} + (-)^n 4b\kappa \sin\phi \tag{47a}$$

$$\int_0^{\pi/2} \left[ e^{-2b\kappa\sec^2\theta} J_{2n+1}(b\kappa \sec^2\theta \sin\theta) \cos(b\kappa \cos\phi \sec^2\theta \sin\theta) \frac{\sec^3\theta}{\sin\theta} \right] d\theta$$

For symmetrical wing we require that both the chord $c(y)$ and the incidence $\alpha_a(y)$ are even in $y$, hence we may expand the following quantities into Fourier series with known coefficients

$$\frac{c_0 \sin\phi}{c(y)} = \sum_{n=0}^{\infty} c_{2n} \cos 2n\phi, \tag{48}$$

$$\alpha_a(y) \sin\phi = \sum_{n=0}^{\infty} B_{2n} \cos 2n\phi. \tag{49}$$

Substituting Eqs. (45)–(49) into Eq. (42), we get

$$\frac{8b}{c_0 a_s} \left[ \sum_{n=0}^{\infty} c_{2n} \cos 2n\phi \right] \left[ \sum_{n=0}^{\infty} A_{2n+1} \sin(2n+1)\phi \right] = \left[ \sum_{n=0}^{\infty} B_{2n} \cos 2n\phi \right] \left[ 1 - \sum_{n=0}^{\infty} (-)^n (2n+1) A_{2n+1} \mathfrak{U}_{2n+1}(\phi) \right] - \sum_{n=0}^{\infty} (2n+1) A_{2n+1} \mathfrak{W}_{2n+1}(\phi) \tag{50}$$

where $\mathfrak{U}_{2n+1}$ and $\mathfrak{W}_{2n+1}$ are given by Eqs. (46a) and (47a). To determine the coefficients $A_{2n+1}$ from Eq. (50) would necessitate the expansion of the left side, and both $\mathfrak{U}_{2n+1}(\phi)$ and $\mathfrak{W}_{2n+1}(\phi)$ into a Fourier series in $\phi$, thus leading to infinitely many linear equations, in infinitely many unknowns. The solution to this problem is obviously a difficult one. However, a very good approximation can be obtained by using a method for practical calculation on wings of finite span, due to H. Glauert[2]. Replace the infinite series representation of $\Gamma(y)$ given by Eq. (45) by an approximate value in terms of a finite series of, say, $m+1$ terms,

$$\Gamma(y) = 4bU \sum_{n=0}^{m} A_{2n+1} \sin(2n+1)\phi, \qquad y = b\cos\phi. \tag{51}$$

This expression of $\Gamma(y)$ then reduces the condition of Eq. (42) to

$$\frac{8b}{a_s} \left[ \frac{\sin\phi}{c(y)} \right] \left[ \sum_{n=0}^{m} A_{2n+1} \sin(2n+1)\phi \right] = (\alpha_a(y) \sin\phi) \cdot \left[ 1 - \sum_{n=0}^{m} (-)^n (2n+1) A_{2n+1} \mathfrak{U}_{2n+1}(\phi) \right] - \sum_{n=0}^{m} (2n+1) A_{2n+1} \mathfrak{W}_{2n+1}(\phi). \tag{52}$$

This equation cannot be satisfied identically for all values of $\phi$. However, if $(m + 1)$ particular values are suitably chosen for $\phi$, we get $m + 1$ linear equations from which the $(m + 1)$ coefficients, $A_1, A_3, \cdots A_{2m+1}$ can be determined. The values of these $A_{2n+1}$ so obtained will satisfy Eq. (52), not identically, but only at the selected points. In general, the first four coefficients usually give a sufficiently accurate result. The detail of such calculations will not be given here; however, the calculation of the first order term is quite similar to that of a specific example of the direct problem discussed in the section "Example—Elliptical Distribution of the Circulation Strength."

Having obtained the value of $\Gamma(y)$ for this indirect problem, the calculation of the lift and the drag is then the same as that of the direct problem discussed in section "Calculation of the Drag from Induced Velocities" and section "Calculation of the Total Lift." The results can be directly written down by ascribing to $f(\mu)$ and $g(\mu)$ the following value:

$$f(\mu) = \frac{1}{\pi}\int_{-b}^{b} \Gamma(\eta)\cos\mu\eta\, d\eta = -\frac{1}{\pi\mu}\int_{-b}^{b} \sin\mu\eta \frac{d\Gamma(\eta)}{d\eta}\, d\eta$$

$$= \frac{4bU}{\pi\mu}\sum_{n=0}^{\infty}(2n+1)A_{2n+1}\int_0^{\pi}\sin(\mu b\cos\psi)\cos(2n+1)\psi\, d\psi \qquad (53a)$$

$$= \frac{4bU}{\mu}\sum_{n=0}^{\infty}(-)^n(2n+1)A_{2n+1}J_{2n+1}(b\mu)$$

and

$$g(\mu) = 0 \qquad (53b)$$

**Minimum Drag for Given Lift.** We shall consider a hydrofoil of span 2b so that $\Gamma(y) = 0$ for $|y| \geqq b$. The problem is to find a distribution $\Gamma(y)$ such that the drag $D$ is minimum under the condition of constant lift $L$. By the method of undetermined multipliers the requirement is

$$\delta D - \lambda_0 \delta L = 0 \qquad (54)$$

where the variation is carried out such that $\delta\Gamma(y) = 0$ at $y = \pm b$. Applying the above variation to the general expressions for $D$ and $L$ given by Eqs. (26a)–(29b), (31), (35a), and (35b), using the relation that $\delta[\Gamma(\eta)\Gamma(\eta')] = \Gamma(\eta)\delta\Gamma(\eta') + \Gamma(\eta')\delta\Gamma(\eta)$ and noting that both $F(\eta - \eta')$ and $G(\eta - \eta')$ defined respectively by Eqs. (35b) and (29b) are symmetric with respect to $(\eta - \eta')$, we obtain

$$\int_{-b}^{b}\left[-2\frac{\partial\varphi(0, y, -h)}{\partial z} - \lambda_0\left(U + 2\frac{\partial\varphi(0, y, -h)}{\partial x}\right)\right]\delta\Gamma(y)\, dy = 0.$$

Since $\delta\Gamma(y)$ is otherwise arbitrary, the quantity inside the bracket must vanish identically. Or, using the present notation, we have

$$2w(y) - \lambda_0[U + 2u(y)] = 0. \qquad (55)$$

The constant multiplier $\lambda_0$ determined by Eq. (55) equals approximately twice

the value of the downwash angle $\epsilon$ for not too shallow submergences; more precisely, the difference $\epsilon - \lambda_0/2$ is a second order small quantity given by

$$\epsilon - \frac{\lambda_0}{2} = \frac{u(y)w(y)}{[U + u(y)][U + 2u(y)]}.$$

Therefore, the condition for the minimum total drag, accounting for all causes and holding lift constant, is that the total downwash angle must be constant, along the span, up to the first order term. When depth of submergence becomes infinite, $u(y)$ tends to zero, the above condition is reduced to the requirement of constant downwash which is in agreement with aerodynamic wing theory.

Substituting the explicit expressions of the induced velocities $w(y)$ and $u(y)$ given by Eqs. (43) and (44) into Eq. (55), we obtain the following integral equation for $\Gamma(y)$:

$$\frac{1}{2\pi}\int_{-b}^{b}\Gamma(\eta)\,d\eta\left\{\int_0^\infty (1 - e^{-2h\mu})\cos\mu(y-\eta)\mu\,d\mu + 4\kappa^2 \cdot\int_0^{\pi/2} e^{-2\kappa h\sec^2\theta}\cos(\kappa(y-\eta)\sec^2\theta\sin\theta)\sec^5\theta\,d\theta\right\} = \lambda_0 \tag{55a}$$

$$\cdot\left\{U - \frac{1}{\pi^2}\int_{-b}^{b}\Gamma(\eta)\,d\eta\int_0^\infty e^{-2h\mu}\mu\,d\mu\int_0^{\pi/2}\cos(\mu(y-\eta)\sin\theta)\frac{\mu + \kappa\sec^2\theta}{\mu - \kappa\sec^2\theta}d\theta\right\}$$

To solve this equation a practical method of approximation similar to that used in solving Eq. (42) of the indirect problem may also be applied here. Assuming that the exact solution of Eq. (55a) may be approximated by the expression given by Eq. (51), Eq. (55) can be reduced to the form

$$\sum_{n=0}^{m}(2n+1)A_{2n+1}\mathfrak{W}_{2n+1}(\phi) = \lambda_0\sin\phi[\tfrac{1}{2} - \sum_{n=0}^{m}(-)^n(2n+1)A_{2n+1}\mathfrak{U}_{2n+1}(\phi)] \tag{56}$$

where $\mathfrak{U}_{2n+1}$ and $\mathfrak{W}_{2n+1}$ are given by Eqs. (46a) and (47a) respectively. From this equation the $(m+1)$ coefficients $A$ may then be determined for $(m+1)$ particularly selected values of $\phi$. It should be remarked here again that the solution so obtained cannot satisfy Eq. (56) identically in $\phi$, but nevertheless gives a good approximation. The estimation of the deviation of this solution from the elliptical distribution, which is the solution of the corresponding problem in aerodynamic wing theory, will be made in the specific example discussed below.

**Example—Elliptical Distribution of the Circulation Strength.** As shown in the indirect problem that the first term in the assumed expansion of $\Gamma(y)$ given by Eq. (45) represents the so-called elliptical distribution, it would be of interest to consider a direct problem of a hydrofoil of span $2b$ immersed at a fixed depth $h$ with an elliptical distribution of $\Gamma(y)$, that is,

$$\begin{aligned}\Gamma(y) &= \Gamma_0\sqrt{1 - y^2/b^2}, && |y| \leqq b;\\ &= 0, && |y| \geqq b.\end{aligned} \tag{57}$$

Although for a given hydrofoil of fixed geometrical form this distribution cannot be held for different values of $h$, nevertheless it gives interesting results. In this problem we shall further define the Froude number of the motion by

$$\sigma = 1/h\kappa = U^2/gh \tag{58a}$$

which is one of important parameters of this problem. We shall mostly be concerned with large values of $\sigma$, corresponding to shallow submergences and large $U$. For instance, for $U$ equal to 80 ft/sec and $h$ equal to 4 ft, $\sigma$ is approximately 50. We shall further denote

$$\beta = 1/b\kappa = U^2/gb, \qquad \beta/\sigma = h/b = \lambda. \tag{58b}$$

$\beta$ is in general also very large in our velocity range of interest.

For this distribution of $\Gamma(y)$ we find, from Eq. (27),

$$f(\mu) = \frac{2\Gamma_0}{\pi}\int_0^b \sqrt{1-\frac{y^2}{b^2}}\cos(\mu y)\,dy = \Gamma_0\frac{J_1(\mu b)}{\mu}, \qquad g(\mu) = 0. \tag{59}$$

a. *Calculation of the Drag.* Substituting Eq. (59) into Eq. (26), we have (cf. Ref. 25, p. 405):

$$D_1 = \frac{\pi\rho\Gamma_0^2}{4}\int_0^\infty \frac{J_1^2(\mu b)}{\mu}\,d\mu = \frac{\pi\rho\Gamma_0^2}{8}; \tag{60}$$

and from Eq. (28) we obtain (cf. Ref. 1)

$$\begin{aligned} D_2 + D_3 &= -\frac{\pi\rho\Gamma_0^2}{4}\int_0^\infty e^{-2h\mu}\frac{J_1^2(\mu b)}{\mu}\,d\mu \\ &= -\frac{\pi\rho\Gamma_0^2}{8}\left\{1-\frac{4}{\pi}\lambda\sqrt{1+\lambda^2}\left[K\left(\frac{1}{\sqrt{1+\lambda^2}}\right)-E\left(\frac{1}{\sqrt{1+\lambda^2}}\right)\right]\right\} \end{aligned} \tag{61}$$

where $K(k)$ and $E(k)$ denote the complete elliptic integral of the first and second kind respectively. Using the series expansions and the asymptotic expansions of $K$ and $E$ (cf. Ref. 27, p. 73), we find the behavior of $(D_2 + D_3)$ for small and large values of $\lambda$ respectively as follows:

$$\begin{aligned} D_2 + D_3 \cong -\frac{\pi\rho\Gamma_0^2}{8}\Bigg\{1-\frac{4}{\pi}\frac{\lambda}{\sqrt{1+\lambda^2}}\Bigg[\log\frac{4\sqrt{1+\lambda^2}}{\lambda} - 1 + \frac{3}{4}\frac{\lambda^2}{(1+\lambda^2)} \\ \left(\log\frac{4\sqrt{1+\lambda^2}}{\lambda}-\frac{4}{3}\right) + 0\,(\lambda^2\log\lambda)\Bigg]\Bigg\}, \qquad \text{as } \lambda \to 0; \end{aligned} \tag{62a}$$

$$D_2 + D_3 = -\frac{\pi\rho\Gamma_0^2}{64\lambda^2}\left\{1-\frac{1}{4\lambda^2}+0\left(\frac{1}{\lambda^4}\right)\right\}, \qquad \text{as } \lambda \to \infty. \tag{62b}$$

As $h \to \infty$, the drag due to the surface effect diminishes like $1/h^3$. When the hydrofoil becomes a planning surface, i.e. $h \to 0$, we find that $(D_2 + D_3)$ tends to a finite value $-\pi\rho\Gamma_0^2/8$ which cancels $D_1$, and consequently, the drag is then solely due to the wave effect.

To find the wave drag, we substitute Eq. (59) into Eq. (29) and obtain the following integral*:

$$D_4 = \pi\rho\Gamma_0^2 \int_0^{\pi/2} e^{-(2/\sigma)\sec^2\theta} \left[\frac{J_1\left(\frac{1}{\beta}\sec^2\theta \sin\theta\right)}{\sec^2\theta \sin\theta}\right]^2 \sec^5\theta \, d\theta. \tag{63}$$

For general values of the Froude number $\sigma$, the above integral may be converted into an infinite series, each term of which contains modified Bessel functions of the second kind, (cf. App.), as follows:

$$D_4 = \frac{\rho\Gamma_0^2}{4}\frac{e^{-1/\sigma}}{\beta^2}\sum_{n=0}^{\infty}\frac{(-)^n\Gamma\left(n+\frac{1}{2}\right)\Gamma\left(n+\frac{3}{2}\right)}{n!(n+1)!(n+2)!}\left(\frac{\sigma}{2\beta^2}\right)^n \left\{K_n\left(\frac{1}{\sigma}\right) + [1+\sigma(n+\tfrac{1}{2})]K_{n+1}\left(\frac{1}{\sigma}\right)\right\}. \tag{64}$$

For very small values of $\sigma$ ($\sim$ large $hU^{-2}$, or $\lambda U^{-2}$), we can use the asymptotic expansion of $K_n(1/\sigma)$,

$$K_n\left(\frac{1}{\sigma}\right) \cong \sqrt{\frac{\pi\sigma}{2}}\, e^{-1/\sigma}\left[1 + \frac{4n^2-1}{1!8}\sigma + \frac{(4n^2-1)(4n^2-3^2)}{2!(8)^2}\sigma^2 + \cdots\right]. \tag{65}$$

So we see that in this case the series in Eq. (64) converges very fast and the final result may be expressed asymptotically by

$$D_4 \cong \frac{\pi\rho\Gamma_0^2}{8}\frac{e^{-2/\sigma}}{\beta^2}\left(\frac{\pi\sigma}{2}\right)^{\frac{1}{2}}\left[1 + \frac{3}{8}\left(1 - \frac{1}{6\beta^2}\right)\sigma + 0(\sigma^2)\right], \qquad \text{as } \sigma \to 0, \tag{66}$$

which diminishes exponentially with increasing $h$. As a remark, the above asymptotic expression of $D_4$ for $\sigma$ small can also be obtained by evaluating the integral of Eq. (63) in a complex $\theta$ plane and applying the method of steepest descent (e.g. cf. Ref. 24, p. 504).

For large values of $\sigma$ ($\sim$ small $h/U^2$), $K_n(1/\sigma)$ has the following expansion:

$$\begin{aligned} K_0\left(\frac{1}{\sigma}\right) &= \log(2\sigma)\cdot I_0\left(\frac{1}{\sigma}\right) + \sum_{m=0}^{\infty}\frac{(2\sigma)^{-2m}}{(m!)^2}\psi(m+1), \\ K_n\left(\frac{1}{\sigma}\right) &= \frac{1}{2}\sum_{m=0}^{n-1}\frac{(-)^m(n-m-1)!}{m!}(2\sigma)^{n-2m} + (-)^n\sum_{m=0}^{\infty}\frac{(2\sigma)^{-n-2m}}{m!(n+m)!} \\ &\quad \cdot[\log(2\sigma) + \tfrac{1}{2}\psi(m+1) + \tfrac{1}{2}\psi(n+m+1)], \quad n = 1, 2, \cdots. \end{aligned} \tag{67}$$

It can be seen that the most important contribution to the sum of the series in Eq. (64) comes from the term $\frac{1}{2}n!(2\sigma)^{n+1}$ of the expansion of $K_{n+1}(1/\sigma)$. As shown

* From the integral representation of the wave drag given by Eq. (63), we note that

$$D_4 = \rho\Gamma_0^2 f(\sigma, \lambda), \qquad \text{or,} \qquad D_4/\rho U^2 C_0^2 = fn(\sigma, \lambda)$$

which can be directly derived from dimensional analysis (cf. Ref. 5, p. 438). This is a useful non-dimensional form for experimental purpose.

in Appendix, this term results in a series which converges only for $\lambda \geqq 1$ and represents, in this region, an analytic function of $\lambda$. Consequently the value of the wave drag $D_4$ for shallow submergences ($\sim\lambda < 1$) may be obtained by some analytical continuation of this function from the region $\lambda \geqq 1$ to the whole region of physical interest, namely, $0 \leqq \lambda < \infty$ (cf. App.). The final result is

$$D_4 \cong \frac{\pi\rho\Gamma_0^2}{4}\left\{1 - \frac{4}{\pi}\lambda\sqrt{1+\lambda^2}\left[K\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - E\left(\frac{1}{\sqrt{1+\lambda^2}}\right)\right]\right.$$
$$+ \frac{4}{3\beta}\left[\frac{2}{\pi}(1+\lambda^2)^{3/2}E\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - \tfrac{3}{2}\lambda - \lambda^3\sqrt{1+\lambda^2}\right. \tag{68}$$
$$\left.\left.\cdot F\left(-\frac{1}{2}, \frac{3}{2}; 1; \frac{1}{1+\lambda^2}\right)\right] + \frac{3}{16\beta^2}\log\beta + 0\left(\frac{1}{\beta^2}, \frac{1}{\sigma}\right)\right\}, \quad \text{as } \sigma \to \infty .$$

It is of interest to note that when the hydrofoil is close to the surface, the wave drag is rather similar in behavior but of the opposite sign and numerically almost twice of the drag due to the surface effect (cf. Eq. (61)). However, it should be emphasized here that even $\Gamma(y)/\Gamma_0$ were the same function of $y$ for all values of $\lambda$, $\Gamma_0$ still depend on $\lambda$, especially for $\lambda$ small, (cf. Eqs. (9) and (41a)). Consequently the actual dependence of the induced and wave drag on $\lambda$ for a given hydrofoil at fixed geometric incidence can only be determined after the function $\Gamma_0(\lambda)$ is found later in this section, part (d).

It is also of interest to compare, when $\sigma$ is small, the first term in Eq. (64),

$$D' = \frac{\pi\rho\Gamma_0^2}{16}\left(\frac{g^2b^2}{U^4}\right)e^{-1/\sigma}\left[K_0\left(\frac{1}{\sigma}\right) + \left(1+\frac{\sigma}{2}\right)K_1\left(\frac{1}{\sigma}\right)\right],$$

with the wave resistance on a moving sphere of radius $r$ given approximately by (cf. Ref. 6, p. 437, also Ref. 7):

$$R = \pi\rho\left(\frac{g^2r^3}{U^3}\right)^2 e^{-1/\sigma}\left[K_0\left(\frac{1}{\sigma}\right) + \left(1+\frac{\sigma}{2}\right)K_1\left(\frac{1}{\sigma}\right)\right], \quad \left(\sigma = \frac{U^2}{gh}\right),$$

Except for a proportionality constant which depends on the size of the obstacle, they have the same dependence on $\sigma$. Evidently an observer on water surface cannot distinguish bodies of different forms if they move at sufficiently great depth.

b. *Calculation of the Lift.* The total lift experienced by the hydrofoil can be obtained, by substituting Eq. (59) into Eqs. (31)–(35), as follows:

$$L = L_0 + \Delta L_1 + \Delta L_2 \tag{69}$$

where

$$L_0 = \rho U\Gamma_0\int_{-b}^{b}\sqrt{1 - \frac{y^2}{b^2}}\,dy = \frac{\pi}{2}\rho Ub\Gamma_0, \tag{70}$$

$$\Delta L_1 = -\frac{\rho\Gamma_0^2}{2}\int_0^{\pi/2}\frac{d\theta}{\sin^2\theta}\int_0^{\infty} e^{-2\lambda\mu}J_1^2(b\mu\sin\theta)\,\mu^{-1}\,d\mu, \tag{71}$$

$$\Delta L_2 = -\kappa\rho\Gamma_0^2\int_0^{\pi/2}\frac{d\theta}{\sin^2\theta}\int_0^{\infty} e^{-2\lambda\mu}J_1^2(b\mu\sin\theta)\,\frac{\sec^2\theta}{\mu(\mu - \kappa\sec^2\theta)}\,d\mu. \tag{72}$$

It can be seen that $L_0$ is the aerodynamic value of the lift, $\Delta L_1$ is independent of the gravitational effect and thus represents the correction due to the mean free surface. All the lift caused by the surface wave effects is represented by $\Delta L_2$. For small incidence angles, $\Gamma_0$ is approximately proportional to the incidence angle (cf. Eq. (41a)), hence relative to $L_0$, $\Delta L$ and $D$ are second order small quantities.

The integral representing $\Delta L_1$ contains the same integral given in Eq. (61) (cf. also Ref. 1), the result of which may be then applied here. After this substitution, integrating by parts with respect to $\theta$, we obtain a simpler representation for $\Delta L_1$ (cf. Ref. 1):

$$\Delta L_1 = -\frac{\rho\Gamma_0^2}{\pi\lambda}\int_0^{\chi}\left(\frac{1-\left(\frac{k}{\chi}\right)^2}{1-k^2}\right)^{\frac{1}{2}} C(k)\,dk \tag{73}$$

where $C(k)$ is a derived complete elliptic integral (cf. Ref. 27, p. 73) defined by

$$C(k) = \int_0^{\pi/2} \frac{\sin^2\phi\cos^2\phi}{[1-k^2\sin^2\phi]^{3/2}}\,d\phi \tag{73a}$$

and

$$\chi = (1+\lambda^2)^{-\frac{1}{2}} \tag{73b}$$

For $0 < h < \infty$ we have the corresponding values of $\lambda$ and $\chi$ as $0 < \lambda < \infty$ and $1 > \chi > 0$. When $\lambda \to 0$ ($\chi \to 1$), the integral in Eq. (73) tends to a definite value, namely

$$\lim_{\chi\to 1-0}\int_0^{\chi}\left[\frac{1-(k/\chi)^2}{1-k^2}\right]^{\frac{1}{2}} C(k)\,dk = \int_0^1 C(k)\,dk = \tfrac{1}{3}[kB(k)-(1-k^2)kC(k)]_0^1 = \tfrac{1}{3} \tag{74}$$

where $B(k)$ is another derived complete elliptic integral (cf. Ref. 27, p. 78) defined by

$$B(k) = (1/k^2)\,[E(k) - (1-k^2)K(k)]. \tag{74a}$$

Hence Eq. (73) indicates that $\Delta L_1$ tends to $\infty$ as $\lambda \to 0$ noting that $\Gamma_0^2$ does not vanish as fast as $\lambda$, as will be shown later. This fact is not surprising because it is known that as the hydrofoil approaches the planning condition, the total lift drops to almost half of its aerodynamic value $L_0$ (e.g. cf. Ref. 15). The divergence of the integral in Eq. (71) simply implies that this second order small quantity will grow so large that it will modify the value of the first order quantity $L_0$.

For $\lambda$ not too small ($\sim\chi$ not too close to 1), the integral in Eq. (73) can be calculated by using the known expansion of $C(k)$, (cf. Ref. 27, p. 73), given as follows:

$$C(k) = \frac{\pi}{16}\left[1 + 6\left(\frac{k^2}{8}\right) + \frac{75}{2}\left(\frac{k^2}{8}\right)^2 + 245\left(\frac{k^2}{8}\right)^3 + 0(k^8)\right]. \tag{75}$$

Substituting this expansion into Eq. (73), then applying the transformation $k^2 = \chi^2 t$ and integrating termwise, we obtain:

$$\Delta L_1 = -\frac{\pi\rho\Gamma_0^2}{64\lambda\sqrt{1+\lambda^2}}\left[1 + \frac{5}{16}\frac{1}{(1+\lambda^2)} + \frac{181}{512}\frac{1}{(1+\lambda^2)^2} + \frac{3015}{8^3}\frac{1}{(1+\lambda^2)^3} + 0\left(\frac{1}{(1+\lambda^2)^4}\right)\right]. \tag{76}$$

This formula should be good for $\lambda > 1$ (or $h > b$). For $\chi$ close to 1 ($h \approx 0$), the above series converges slowly; hence we have to resort to some other approximation in order to facilitate calculation. For this case, we define

$$\chi'^2 = 1 - \chi^2, \qquad \chi'^2 \ll 1, \qquad \chi'^2 = \frac{\lambda^2}{1+\lambda^2} = \lambda^2 + 0(\lambda^4). \tag{77}$$

Then

$$\left[\frac{1 - (k/\chi)^2}{1 - k^2}\right]^{\frac{1}{2}} = 1 - \tfrac{1}{2}\chi'^2\left(\frac{k^2}{1-k^2}\right) + 0(\chi'^4);$$

and

$$\Delta L_1 = -\frac{\rho\Gamma_0^2}{\pi\lambda}\left[\int_0^\chi C(k)\,dk - \tfrac{1}{2}\chi'^2\int_0^\chi \frac{k^2}{1-k^2}C(k)\,dk + 0(\chi'^4)\right].$$

Now the first integral can be integrated,

$$\int_0^\chi C(k)\,dk = \tfrac{1}{3}[\chi B(\chi) - \chi\chi'^2 C(\chi)],$$

and the second integral may be approximated by

$$\int_0^\chi \frac{k^2}{1-k^2}C(k)\,dk = -\tfrac{1}{4}\log\sqrt{1-\chi^2} = -\tfrac{1}{4}\log\frac{\lambda}{\sqrt{1+\lambda^2}}.$$

Finally we have

$$\Delta L_1 \cong -\frac{\rho\Gamma_0^2}{3\pi\lambda}\left\{\frac{1}{\sqrt{1+\lambda^2}}B\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - \frac{\lambda^2}{(1+\lambda^2)}\cdot\left[\frac{1}{\sqrt{1+\lambda^2}}C\left(\frac{1}{\sqrt{1+\lambda^2}}\right) + \frac{3}{8}\log\frac{\sqrt{1+\lambda^2}}{\lambda}\right] + 0(\lambda^4\log\lambda)\right\}, \tag{78}$$

and using the known expansion of $B(\chi)$ and $C(\chi)$ as $\chi \to 1$ we obtain

$$\Delta L_1 \cong -\frac{\rho\Gamma_0^2}{3\pi\lambda}\left\{1 - 3\lambda^2\left[\log 2 - \frac{3}{4} + \frac{5}{8}\log\frac{\sqrt{1+\lambda^2}}{\lambda}\right] + 0(\lambda^4\log\lambda)\right\}, \tag{78a}$$

as $\lambda \to 0$.

The integral representing $\Delta L_2$ in Eq. (72) is indeterminate, but its principle value exists. The method of contour integration described previously in the section "General Formulation of the Problem" gives a complicated expression. In

this case we shall use the following approximation. By letting $\mu = (\kappa \sec^2\theta)u$ we can transform Eq. (72) into the following form:

$$\Delta L_2 = -\rho\Gamma_0^2 \int_0^\infty \frac{du}{u(u-1)} \int_0^{\pi/2} e^{-(2u/\sigma)\sec^2\theta} \left[\frac{J_1\left(\frac{u}{\beta}\sec^2\theta\sin\theta\right)}{\sec^2\theta\sin\theta}\right]^2 \sec^4\theta\, d\theta \qquad (79)$$

The inner integral is very similar to that representing $D_4$ in Eq. (63) which is discussed in detail in Appendix. This integral may be treated in a similar way. Because the interesting feature of $\Delta L_2$ also comes from the integration with respect to $u$, we shall take only the first order term of the inner integral. The final result is as follows (cf. Ref. 1): for $\sigma$ very large (shallow submergence),

$$\Delta L_2 = \frac{\rho\Gamma_0^2}{4\sqrt{2}}\Gamma(1/4)\left\{1 - \frac{\Gamma(3/4)}{\sqrt{\pi}\Gamma(1/4)}\frac{\lambda}{\sqrt{1+\lambda^2}}\right\}\left[1 + \sqrt{\frac{2}{\pi\sigma}}\left(\gamma + \log\frac{2}{\sigma}\right)\right] \left(1 + 0\left(\lambda^2, \frac{1}{\sigma}\right)\right), \qquad (80)$$

where $\gamma$ is Euler's constant, $\gamma = 0.5772$; and for $\sigma$ very small, the inner integral in Eq. (79) can first be approximated by using the method of steepest descent, so that the asymptotic value of $\Delta L_2$ is

$$\Delta L_2 \cong -\frac{\rho\Gamma_0^2}{8}\sqrt{\frac{\pi\sigma}{2}}\frac{1}{\beta^2}\int_0^\infty e^{-2u/\sigma}\frac{\sqrt{u}}{u-1}\,du\,(1 + 0(\sigma)).$$

The principal value of the above integral for $\sigma$ small may be then approximated by using the generalized Watson's lemma (cf. Ref. 24, p. 502) to yield:

$$\Delta L_2 \cong \frac{\pi\rho\Gamma_0^2}{64\lambda^2}[1 + 3/4\sigma + 15/16\sigma^2 + 0(\sigma^3)]. \qquad (81)$$

From these relations it is of interest to see that $\Delta L_2$ tends to a value

$$\rho\Gamma_0^2\Gamma(1/4)/4\sqrt{2}$$

as $h$ tends to zero. This result is quite a contrast to the feature of $\Delta L_1$. Compared with $\Delta L_1$, the value of $\Delta L_2$ is smaller for $\sigma$ large, and greater for $\sigma$ small.

c. *The Induced Velocity at the Hydrofoil.* In order to investigate further the geometric and hydrodynamic properties of the hydrofoil, we need to know the value of $\Gamma_0$; and if we want to calculate $\Gamma_0$ in terms of given quantities, we have first to obtain the values of the $u$-velocity and the downwash at the hydrofoil (cf. Eq. 41).

Substituting the value of $\Gamma(y)$ into Eq. (43), we may express $u(y)$ in terms of two parts:

$$u(y) = u_1(y) + u_2(y) \qquad (82)$$

where

$$u_1(y) = -\frac{\Gamma_0}{2\pi b}\int_0^{\pi/2} d\theta \int_0^\infty e^{-2\lambda\mu}\frac{J_1(\mu\sin\theta)}{\sin\theta}\cos(\mu\eta\sin\theta)\,d\mu, \qquad (82a)$$

$$u_2(y) = -\frac{\kappa\Gamma_0}{\pi}\int_0^{\pi/2} d\theta \int_0^{\infty} e^{-2\lambda\mu}\frac{J_1(\mu\sin\theta)}{\sin\theta}\cos(\mu\eta\sin\theta)\frac{\sec^2\theta}{\mu - b\kappa\sec^2\theta}\,d\mu, \quad (82b)$$

and

$$\eta = y/b, \qquad \kappa = g/U^2, \qquad \lambda = h/b. \quad (82c)$$

The integral in Eq. (82a), which represents the effect due to the mean free surface, can be evaluated exactly. Expanding both $J_1(\mu\sin\theta)$ and $\cos(\mu\eta\sin\theta)$ into power series, then integrating termwise and summing up the series which is resulted from $J_1$ expansion, we obtain

$$u_1(y) = -\frac{\Gamma_0}{8\sqrt{\pi}b}\frac{1}{\lambda\sqrt{1+4\lambda^2}}\sum_{n=0}^{\infty}\frac{(-)^n\Gamma(n+3/2)}{n!}\frac{\eta^{2n}}{(1+4\lambda^2)^n} \cdot {}_2F_1\left(n+\tfrac{1}{2}, \tfrac{1}{2}-n; 2; \frac{1}{1+4\lambda^2}\right). \quad (83)$$

The value of $u_1(y)$ is always negative and becomes infinite as $h \to 0$. It can be shown that the above series converges uniformly with respect to $\eta$ when $0 \leqq \eta \leqq 1$. We may also find that $u_1(y)$ is a slowly varying function of $y$; in particular, the value of $u_1(0)$ is

$$u_1(0) = -\frac{\Gamma_0}{16b}\frac{1}{\lambda\sqrt{1+4\lambda^2}}\,{}_2F_1\left(\tfrac{1}{2}, \tfrac{1}{2}; 2; \frac{1}{1+4\lambda^2}\right) = -\frac{\Gamma_0}{4\pi b}\frac{1}{\lambda\sqrt{1+4\lambda^2}}B\left(\frac{1}{\sqrt{1+4\lambda^2}}\right) \quad (84)$$

where $B(k)$ is a derived complete elliptic integral defined by Eq. (74a).

It can be shown that $u_2(y)$ is also a slowly varying function of $y$, i.e. the values of $u_2(y)$ at points close to the hydrofoil differ very slightly from $u_2(0)$, its value at plane of symmetry. Applying the transformation $\mu = (b\kappa\sec^2\theta)u$ to Eq. (82b), we have

$$u_2(0) = -\frac{\kappa\Gamma_0}{\pi}\int_0^{\infty}\frac{du}{u-1}\int_0^{\pi/2} e^{-(2u/\sigma)\sec^2\theta}\frac{J_1\left(\frac{u}{\beta}\sec^2\theta\sin\theta\right)}{\sec^2\theta\sin\theta}\sec^4\theta\,d\theta.$$

For $\sigma$ very large, the inner integral may be approximated in a manner similar to that described in appendix. The final result is

$$u_2(0) \cong -\frac{\Gamma_0}{4b}\frac{1}{\sqrt{2\lambda}}(1+4\lambda^2)^{-3/4}F\left(\tfrac{3}{4}, \tfrac{3}{4}; 2; \frac{1}{1+4\lambda^2}\right) \cdot\left[1+\sqrt{\frac{2}{\pi\sigma}}\left(\gamma+\log\frac{2}{\sigma}\right)\right]\left(1+0\left(\frac{1}{\sigma}\right)\right)$$

where $\gamma = 0.5772$. The above hypergeometric function can be expressed in terms of tabulated elliptic integrals by using Kummer's transformation:

$$F\left(\tfrac{3}{4}, \tfrac{3}{4}; 2; \frac{1}{1+4\lambda^2}\right) = F\left(\tfrac{3}{2}, \tfrac{3}{2}; 2; \frac{1}{2}-\frac{\lambda}{\sqrt{1+4\lambda^2}}\right) = \frac{8}{\pi}\frac{\sqrt{1+4\lambda^2}}{2\lambda+\sqrt{1+4\lambda^2}}B\left(\sqrt{\frac{1}{2}-\frac{\lambda}{\sqrt{1+4\lambda^2}}}\right)$$

where $B(k)$ is defined by Eq. (74a). Therefore, we obtain finally:

$$u_2(0) \cong -\frac{2\Gamma_0}{\pi b}\frac{1}{\sqrt{2\lambda}}\frac{B(\sqrt{\tfrac{1}{2}-\lambda(1+4\lambda^2)^{-1/2}})}{(1+4\lambda^2)^{1/4}[2\lambda+\sqrt{1+4\lambda^2}]} \cdot\left\{1+\sqrt{\frac{2}{\pi\sigma}}\left(\gamma+\log\frac{2}{\sigma}\right)\right\}\left[1+0\left(\frac{1}{\sigma}\right)\right]. \tag{85a}$$

For $\sigma$ very small, we apply the method of steepest descent to the inner integral and Watson's lemma to the integral with respect to $u$, we have

$$u_2(0) \cong -\frac{\Gamma_0}{32b\lambda^2}\{1+\tfrac{3}{4}\sigma+\tfrac{15}{16}\sigma^2+0(\sigma^3)\}. \tag{85b}$$

It is of interest to see that $u_1(0)$ and $u_2(0)$ are of the same order of magnitude. Therefore the total $u$-component velocity at $y = 0$ is

$$u(0) \cong -\frac{\Gamma_0}{\pi b}\left\{\frac{B((1+4\lambda^2)^{-1/2})}{4\lambda\sqrt{1+4\lambda^2}}+\frac{2B(\sqrt{\tfrac{1}{2}-\lambda(1+4\lambda^2)^{-1/2}})}{\sqrt{2\lambda}(1+4\lambda^2)^{1/4}[2\lambda+\sqrt{1+4\lambda^2}]} \cdot\left[1+\sqrt{\frac{2}{\pi\sigma}}\left(\gamma+\log\frac{2}{\sigma}\right)\right]\left(1+0\left(\frac{1}{\sigma}\right)\right)\right\}, \quad \text{for } \sigma \text{ large;} \tag{86a}$$

and

$$u(0) \cong -\frac{\Gamma_0}{4\pi b}\left\{\frac{1}{\lambda\sqrt{1+4\lambda^2}}B\left(\frac{1}{\sqrt{1+4\lambda^2}}\right)-\frac{\pi}{8\lambda^2} [1+\tfrac{3}{4}\sigma+\tfrac{15}{16}\sigma^2+0(\sigma^3)]\right\}, \quad \text{for } \sigma \text{ small.} \tag{86b}$$

The downwash at the hydrofoil can be calculated in a similar way:

$$w(y) = -\varphi_z(0, y, -h) = -(\varphi_{1_z}+\varphi_{2_z}+\varphi_{3_z}+\varphi_{4_z})_{(0,y,-h)} \tag{87a}$$

where

$$\varphi_{1_z}(0, y, -h) = -\frac{\Gamma_0}{4}\int_0^\infty J_1(\mu b)\cos(\mu y)\,d\mu = \begin{cases} -\dfrac{\Gamma_0}{4b}, & |y| \leqq b; \\ -\dfrac{\Gamma_0}{4b}\left(1-\dfrac{|y|}{\sqrt{y^2-b^2}}\right), & |y| > b, \end{cases} \tag{87b}$$

$$(\varphi_{2_z}+\varphi_{3_z})_{(0,y,-h)} = \frac{\Gamma_0}{4}\int_0^\infty e^{-2h\mu}J_1(\mu b)\cos(\mu y)\,d\mu, \tag{87c}$$

and

$$\varphi_{4_z}(0, y, -h) = -\kappa\Gamma_0\int_0^{\pi/2} e^{-2h\kappa\sec^2\theta}\frac{J_1(\kappa b\sec^2\theta\sin\theta)}{\sec^2\theta\sin\theta} \cdot\cos(\kappa y\sec^2\theta\sin\theta)\sec^5\theta\,d\theta. \tag{87d}$$

Now the integral in Eq. (87c) can be evaluated while that in Eq. (87d) again has to be approximated.

$$\frac{\partial}{\partial z}(\varphi_2 + \varphi_3)_{(0,y,-h)} = \frac{\Gamma_0}{4b}\left\{1 - Rl\frac{2h - i\,|\,y\,|}{\sqrt{(2h - i\,|\,y\,|)^2 + b^2}}\right\} \tag{88a}$$

$$= \frac{\Gamma_0}{4b}\left\{1 - \frac{1}{\sqrt{2}}F_1(\lambda, \eta)\right\}, \qquad \text{(cf. Ref. 26, p. 33)}$$

where

$$F_1(\lambda, \eta) = 2\lambda\left\{\frac{[(4\lambda^2 + 1 - \eta^2)^2 + (4\lambda\eta)^2]^{\frac{1}{2}} + (4\lambda^2 + 1 - \eta^2)}{(4\lambda^2 + 1 - \eta^2)^2 + (4\lambda\eta)^2}\right\}^{\frac{1}{2}} + |\,\eta\,|\left\{\frac{[(4\lambda^2 + 1 - \eta^2)^2 + (4\lambda\eta)^2]^{\frac{1}{2}} - (4\lambda^2 + 1 - \eta^2)}{(4\lambda^2 + 1 - \eta^2)^2 + (4\lambda\eta)^2}\right\}^{\frac{1}{2}} \tag{88b}$$

and

$$\lambda = h/b, \qquad \eta = y/b \tag{88c}$$

It can be seen that $F_1(\lambda, \eta)$ is a slowly varying function of $\eta$ for $\eta < 1$ and for fixed $\lambda$. In particular, we have

$$F_1(\lambda, 0) = \frac{2\sqrt{2}\lambda}{\sqrt{1 + 4\lambda^2}}, \tag{89a}$$

$$F_1(\lambda, \pm 1) = \frac{\lambda\sqrt{\sqrt{1 + \lambda^2} + \lambda} + \sqrt{\sqrt{1 + \lambda^2} - \lambda}}{\sqrt{\lambda(1 + \lambda^2)}}. \tag{89b}$$

For small values of $\sigma$, application of the method of steepest descent to the integral in Eq. (87d) yields

$$\varphi_{4_s}(0, y, -h) \cong -\frac{\Gamma_0}{4b\beta^2}\sqrt{\frac{\pi\sigma}{2}}\, e^{-2/\sigma}(1 + 0(\sigma)) \tag{90}$$

which diminishes exponentially with increasing $h$ and is independent of $y$ for points close to the hydrofoil. For very large $\sigma$, the integral in Eq. (87d) also represents a slowly varying function of $y$ and consequently only the value of $\varphi_{4_s}$ $(0, 0, -h)$ is of interest. It can be shown that (cf. Ref. 1):

$$\varphi_{4_s}(0, 0, -h) \cong -\frac{\Gamma_0}{2b}\left\{1 - \frac{2\lambda}{\sqrt{1 + 4\lambda^2}}\right\}\left[1 + 0\left(\frac{1}{\sigma}, \frac{1}{\beta}\right)\right] \cong -2\frac{\partial}{\partial z}(\varphi_2 + \varphi_3)_{(0,0,-h)}\left[1 + 0\left(\frac{1}{\sigma}, \frac{1}{\beta}\right)\right]. \tag{91}$$

Finally we have the value of $w(0)$, by substituting Eqs. (87a), (88)–(91) into Eq. (87b), as follows

$$w(0) \cong \frac{\Gamma_0}{2b}\left\{1 - \frac{\lambda}{\sqrt{1 + 4\lambda^2}}\right\}\left[1 + 0\left(\frac{1}{\sigma}, \frac{1}{\beta}\right)\right], \qquad \text{for } \sigma \text{ large;} \tag{92a}$$

$$w(0) \cong \frac{\Gamma_0}{2b}\left\{\frac{\lambda}{\sqrt{1 + 4\lambda^2}} + \frac{1}{2\beta^2}\sqrt{\frac{\pi\sigma}{2}}\, e^{-2/\sigma}(1 + 0(\sigma))\right\}, \qquad \text{for } \sigma \text{ small,} \tag{92b}$$

which tends to aerodynamic value

$$w_\infty(0) = \Gamma_\infty(0)/4b \tag{92c}$$

as $h \to \infty$.

It may be remarked here that both $u(y)$ and $w(y)$ are almost constant spanwise at points close to the hydrofoil. This fact justifies some of the basic assumptions introduced previously in the sections, "General Formulation of the Problem" and "Geometry of the Hydrofoil; Effective Angle of Attack." This result also supports the assumption that the same circulation distribution will hold for a wide range of depths.

d. *Geometry of the Hydrofoil.* We are interested primarily in the case of shallow immersion, because the geometry of the hydrofoil at deep immersion will be the same as the corresponding aerodynamic problem. Substituting Eqs. (86) and (92) into (41a), we obtain the value of $\Gamma_0$ ($\equiv \Gamma(0)$) as a function of $\lambda$ and $\sigma$ as follows:

$$\Gamma_0 = \frac{\frac{1}{2}Uc_0 a_e \alpha_a}{1 + (c_0/4b)a_e[f_2(\lambda, \sigma) + (2/\pi)\alpha_a f_1(\lambda, \sigma)]} \tag{93a}$$

where

$$f_1(\lambda, \sigma) = \frac{B((1+4\lambda^2)^{-\frac{1}{2}})}{4\lambda\sqrt{1+4\lambda^2}} + \frac{2B(\sqrt{\frac{1}{2} - \lambda(1+4\lambda^2)^{-\frac{1}{2}}})}{\sqrt{2\lambda}\,(1+4\lambda^2)^{1/4}[2\lambda + \sqrt{1+4\lambda^2}]} \left[1 + \sqrt{\frac{2}{\pi\sigma}}\left(\gamma + \log\frac{2}{\sigma}\right)\right]\left(1 + 0\left(\frac{1}{\sigma}\right)\right), \quad \text{for } \sigma \text{ large;} \tag{93b}$$

$$f_1(\lambda, \sigma) = \frac{1}{4\lambda\sqrt{1+4\lambda^2}} B\left(\frac{1}{\sqrt{1+4\lambda^2}}\right) - \frac{\pi}{32\lambda^2}[1 + \tfrac{3}{4}\sigma + \tfrac{15}{16}\sigma^2 + 0(\sigma^3)], \quad \text{for } \sigma \text{ small,} \tag{93c}$$

and

$$f_2(\lambda, \sigma) = \left\{1 - \frac{\lambda}{\sqrt{1+4\lambda^2}}\right\}\left[1 + 0\left(\frac{1}{\sigma}, \frac{1}{\beta}\right)\right], \quad \text{for } \sigma \text{ large;} \tag{93d}$$

$$f_2(\lambda, \sigma) = \frac{\lambda}{\sqrt{1+4\lambda^2}} + \frac{1}{2\beta^2}\sqrt{\frac{\pi\sigma}{2}}\, e^{-2/\sigma}(1 + 0(\sigma)), \quad \text{for } \sigma \text{ small.} \tag{93e}$$

As $h$ tends to $\infty$, $f_1$ tends to zero and $f_2$ tends to $\frac{1}{2}$; hence the value of $\Gamma_0$ at infinite depth becomes

$$\Gamma_{0_\infty} = \frac{\frac{1}{2}Uc_0 a_e \alpha_a}{1 + \frac{1}{8}(c_0/b)a_e} \tag{94}$$

which is the well-known aerodynamic value of $\Gamma(0)$ in terms of given quantities. It is of interest to note that $\Gamma_0$ changes with respect to $\Gamma_{0_\infty}$, as $\lambda$ and $\sigma$ vary, according to the following relation:

$$\frac{\Gamma_0}{\Gamma_{0_\infty}} = \frac{1 + \frac{1}{8}(c_0/b)a_e}{1 + \frac{1}{4}(c_0/b)a_e[f_2(\lambda, \sigma) + (2/\pi)\alpha_a f_1(\lambda, \sigma)]} = \Psi(\lambda, \sigma, \alpha_a) \tag{95}$$

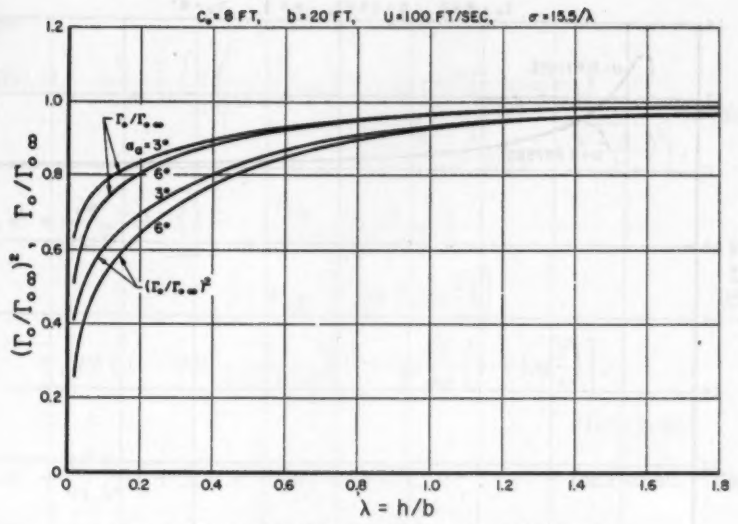



FIG. 3. Effect of depth on circulation strength (large $\sigma$).

which is a function of $\lambda$, $\sigma$ and $\alpha_a$ for a given hydrofoil. This function and its squared value, $\Psi^2$, are plotted against $\lambda$ in Fig. 3 for two incidences $\alpha_a = 3°$ and 6° under the following operating condition:

$$c_0 = 8\,\text{ft}, \quad b = 20\,\text{ft}, \quad a_s \doteq 2\pi, \quad U = 100\,\text{ft/sec}, \quad \sigma = U^2/gh = 15.5/\lambda \qquad (96a)$$

so that

$$\Psi\left(\lambda, \frac{15.5}{\lambda}, \alpha_a\right) = \frac{\Gamma_0}{\Gamma_{0\infty}} = \frac{1.314}{1 + 0.628[f_2(\lambda, 15.5/\lambda) + (2/\pi)\alpha_a f_1(\lambda, 15.5/\lambda)]}. \qquad (96b)$$

The values of $\Psi$ and $\Psi^2$ decrease with decreasing $\lambda$, and tend to unity asymptotically as $\lambda \to \infty$. For small values of $\lambda(\lambda < 1)$, both $\Gamma_0$ and $\Gamma_0^2$ deviate appreciably from their respective values at infinite depth. This result is in agreement with the experimental fact observed by J. S. Ausman[30] that for a two-dimensional hydrofoil moving at large Froude numbers the stagnation point moves from the lower surface across the leading edge to the upper surface of the hydrofoil as it arises from infinite submergence to shallow depths, consequently the circulation developed becomes less and less than that developed at infinite submergence, with the same absolute angle of attack. Although at present there is no other available experimental evidence on such property for three-dimensional hydrofoil, it would still be plausible to believe that a similar behavior would hold for hydrofoils of finite span.

However, the dependence of $\Psi$ on $\lambda$ for small values of $\sigma$ is quite different. For instance, if we take the same hydrofoil as before, moving at a low speed $U =$ 5 ft/sec so that $\sigma = 0.0388/\lambda$ which may be considered small for $\lambda > 0.16$

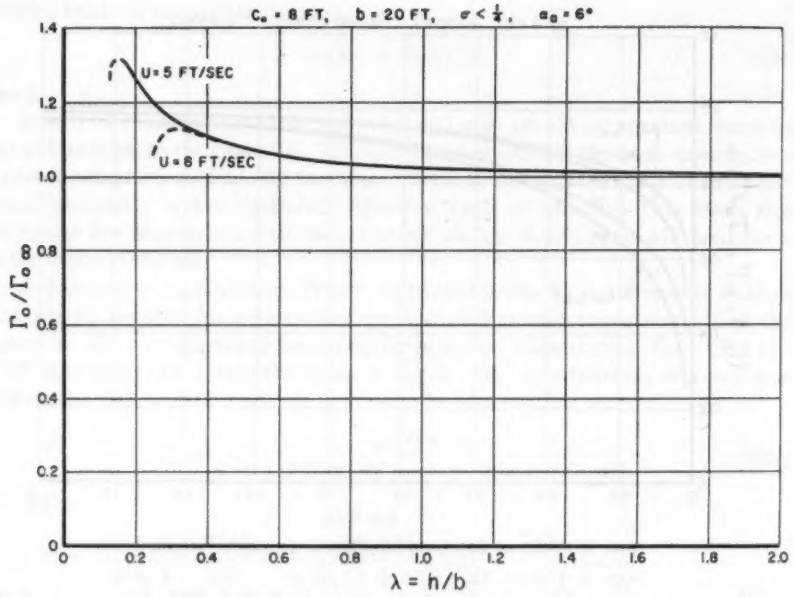



FIG. 4. Effect of depth on circulation strength (small $\sigma$).

($\sigma < \frac{1}{4}$), then the result (cf. Fig. 4) obtained by using Eqs. (93a, c, e) shows that $\Psi$ increases gradually as $\lambda$ decreases from infinity to about 0.15 and then decreases rapidly for further decrease in $\lambda$.

It may be pointed out here that because both $u(y)$ and $w(y)$ are almost constant along the span of the hydrofoil, it follows from Eq. (41) that for an elliptical distribution of $\Gamma(y)$ the plan form is also nearly elliptical,

$$c(y) = c_0\sqrt{1 - y^2/b^2}, \tag{97}$$

together with a negligible geometric twist.

Another consequence of the above results is that the condition of Eq. (55) for minimum drag with prescribed lift is also approximately satisfied by almost constant (spanwise) $u(y)$ and $w(y)$. Hence the hydrofoil of elliptical plan form still gives approximately the minimum drag for a given lift.

e. *Over-all Hydrodynamic Properties; Lift and Drag Coefficient.* In this section we shall again only consider the case of shallow submergence. Combining our previous results we obtain the total lift and drag of the hydrofoil at small depths as follows:

$$L \doteq \frac{\pi}{2}\rho U b \Gamma_{0_\infty}\Psi - \frac{\rho\Gamma_{0_\infty}^2\Psi^2}{3\pi}[f_3(\lambda) - f_4(\lambda, \sigma)] \tag{98a}$$

where

$$f_5(\lambda) = \frac{1}{\lambda\sqrt{1+\lambda^2}}\left\{B\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - \frac{\lambda^2}{1+\lambda^2}C\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - \frac{3}{8}\frac{\lambda^2}{\sqrt{1+\lambda^2}}\log\frac{\sqrt{1+\lambda^2}}{\lambda}\right\}, \tag{98b}$$

$$f_6(\lambda, \sigma) = \frac{3\pi}{4\sqrt{2}}\Gamma(1/4)\left(1 - \frac{\Gamma(3/4)}{\sqrt{\pi}\,\Gamma(1/4)}\frac{\lambda}{\sqrt{1+\lambda^2}}\right)\left[1 + \sqrt{\frac{2}{\pi\sigma}}\left(\gamma + \log\frac{2}{\sigma}\right)\right] \tag{98c}$$

$$= 6.05\left(1 - 0.19\frac{\lambda}{\sqrt{1+\lambda^2}}\right)\left[1 + \sqrt{\frac{2}{\pi\sigma}}\left(\gamma + \log\frac{2}{\sigma}\right)\right],$$

for $\sigma$ large;

$$f_6(\lambda, \sigma) = \frac{3\pi^2}{64}\frac{1}{\lambda^2}[1 + 3/4\sigma + 15/16\sigma^2 + 0(\sigma^3)], \qquad \text{for } \sigma \text{ small,} \tag{98d}$$

and $\Gamma_{0_\infty}$ and $\Psi$ $(\lambda, \sigma, \alpha_a)$ are given by Eqs. (94) and (95) respectively.

$$D \doteq \frac{\pi\rho\Gamma_{0_\infty}^2}{8}\Psi^2\{f_5(\lambda) + f_6(\lambda, \sigma)\}, \tag{99a}$$

where

$$f_5(\lambda) = \frac{4}{\pi}\lambda\sqrt{1+\lambda^2}\left[K\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - E\left(\frac{1}{\sqrt{1+\lambda^2}}\right)\right] \tag{99b}$$

and

$$f_6(\lambda, \sigma) \doteq 2[1 - f_5(\lambda)], \qquad \text{for } \sigma \text{ large;} \tag{99c}$$

$$= \beta^{-2}\sqrt{\pi\sigma/2}e^{-2/\sigma}[1 + 0(\sigma)], \qquad \text{for } \sigma \text{ small.} \tag{99d}$$

At this stage we should be able to see the properties of the total drag exerted on the hydrofoil. As mentioned before, when $\lambda \to \infty$, $D$ tends to the aerodynamic value $D_\infty = \pi\rho\Gamma_{0_\infty}^2/8$, hence we write, for the case of large $\sigma$,

$$D/D_\infty = \Psi^2 f_5(\lambda) + 2\Psi^2[1 - f_5(\lambda)]. \tag{99e}$$

This ratio is plotted in Fig. 5 for $\alpha_a = 6°$ and under the operating condition given by Eq. (96a). The first term on the right hand side represents the contribution from the total induced drag; this quantity tends to $D_\infty$ asymptotically for large $h$ and decreases rapidly as the hydrofoil approaches the surface $(h < 2b)$. The second term represents approximately the wave drag which is negligibly small for $h > 2b$, and become important only at small depths. As a whole, the total drag $D$ deviates only slightly from $D_\infty$ for $h > (b/2)$, and decreases moderately at small depths $(h < (b/2))$ which is actually the range of interest. When

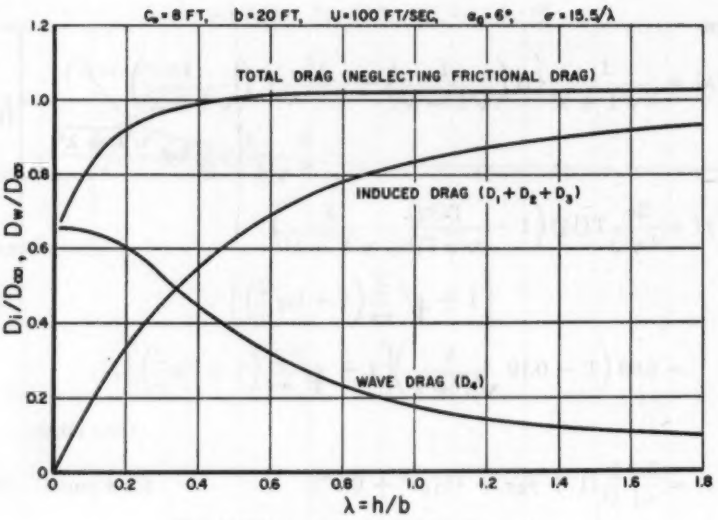



FIG. 5. Effect of depth on drag (large $\sigma$).

the hydrofoil approaches the planning condition, the total drag is solely due to the wave effect.

In order to measure the total drag, including the wave drag, exerted on a hydrofoil, the previous experiments were mostly carried out in a towing tank.* Then it would be convenient to define the over-all lift and drag coefficient with respect to the towing speed, such as,

$$C_L = L/\tfrac{1}{2}\rho U^2 S, \qquad C_D = D/\tfrac{1}{2}\rho U^2 S \tag{100}$$

where $\rho$ is the density of the liquid, $U$ the towing speed, and $S$ the area of the plan form. The above definition for $C_L$ and $C_D$ were used in previous experiments (known to the author). In order to compare with these available experimental results, we shall also adopt the above definition of $C_L$ and $C_D$. Substituting Eqs. (94), (98) and (99) into (100), we obtain

$$C_L \doteq \frac{\pi}{8} A \left(\frac{c_0}{b}\right) \frac{a_e \alpha_0 \Psi}{(1 + \tfrac{1}{8}(c_0/b)a_e)} - \frac{A}{24\pi}\left(\frac{c_0}{b}\right)^2 \left[\frac{a_e \alpha_0 \Psi}{1 + \tfrac{1}{8}(c_0/b)a_e}\right]^2 \cdot \{f_3(\lambda) - f_4(\lambda, \sigma)\}, \tag{101}$$

* The reason for this point is well-known to most experimenters. When the water in a conventional free surface water tunnel of finite depth is driven at sub-critical speeds, the free surface is usually found to have standing waves which are difficult to be eliminated. And if the water is driven at supercritical speeds, the surface appears almost flat without testing model, but, with the model submerged, the feature of the wave drag in deep water is then lost due to the depth effect of the tunnel bottom (cf. Ref. 6, p. 415). However, a new method has been recently introduced by Hoerner and Ausman[20] to eliminate standing waves at subcritical speeds by using an adjustable surface guide vane installed at the contraction throat of the tunnel.

$$C_D \doteq \frac{\pi \mathcal{R}}{64}\left(\frac{c_0}{b}\right)^2\left[\frac{a_e a_a \Psi}{1+\frac{1}{8}(c_0/b)a_e}\right]^2\{f_3(\lambda)+f_6(\lambda,\sigma)\}, \tag{102}$$

where $\mathcal{R} = 4b^2/S$ = aspect ratio and $f_3$, $f_4$, $f_5$, $f_6$ are given in Eqs. (98) and (99). The lift drag ratio is

$$\frac{C_L}{C_D} \doteq \frac{1-\frac{1}{3\pi^2}\left(\frac{c_0}{b}\right)\left[\frac{a_e a_a \Psi}{1+\frac{1}{8}(c_0/b)a_e}\right]\{f_3(\lambda)-f_4(\lambda,\sigma)\}}{\frac{1}{8}\left(\frac{c_0}{b}\right)\left[\frac{a_e a_a \Psi}{1+\frac{1}{8}(c_0/b)a_e}\right]\{f_3(\lambda)+f_6(\lambda,\sigma)\}}. \tag{103}$$

From the expressions of $C_L$ and $C_D$ for $h$ large, we note that as $h \to \infty$, both $C_L$ and $C_D$ tend asymptotically to their aerodynamic values, namely,

$$C_L \cong \frac{\pi}{8}\mathcal{R}\left(\frac{c_0}{b}\right)\frac{a_e a_a}{1+\frac{1}{8}(c_0/b)a_e}, \qquad C_D \cong \frac{\pi}{64}\mathcal{R}\left(\frac{c_0}{b}\right)^2\left[\frac{a_e a_a}{1+\frac{1}{8}(c_0/b)a_e}\right]^2,$$

so that

$$C_D \cong \frac{1}{\pi \mathcal{R}} C_L^2$$

which agrees with aerodynamic wing theory. However, it can also be seen from Eqs. (101) and (102) that even for $\lambda$ small, $C_D/C_L^2$ is still proportional to $\mathcal{R}^{-1}$ if the small quantity of the second order is neglected. From this result it follows that in order to improve the hydrodynamic properties the preference should be for a high aspect ratio hydrofoil, even for operations near the water surface.

To illustrate the details of the behavior, and to compare our results with experiments, we shall take specific values of $c_0$, $b$, $\mathcal{R}$ and $a_e$ to determine all the coefficients. This will be shown as follows.

**Discussion of Results.** Consider a hydrofoil of elliptical plan form with the following specifications

$$c_0 = 8 \text{ ft}, \qquad b = 20 \text{ ft}, \qquad a_e \doteq 2\pi \tag{104a}$$

so that

$$\mathcal{R} = 6.3 \tag{104b}$$

These data determine the coefficients in Eqs. (95), (101)–(103) which then become

$$C_L \doteq 4.74\, a_a\Psi - 0.306\, a_a^2\Psi^2[f_3(\lambda) - f_4(\lambda,\sigma)], \tag{105}$$

$$C_D \doteq 1.14\, a_a^2\Psi^2[f_3(\lambda) + f_6(\lambda,\sigma)] \tag{106}$$

where

$$\Psi = \frac{1.314}{1+0.628[f_2(\lambda,\sigma)+(2a_a/\pi)f_1(\lambda,\sigma)]} \tag{107}$$

and $f_1$ to $f_6$ are given respectively by Eqs. (93), (98), (99) for both large and small values of $\sigma$. In the following, for either large or small $\sigma$, we shall concern only the case of shallow submergence, say, $\lambda < 2$.

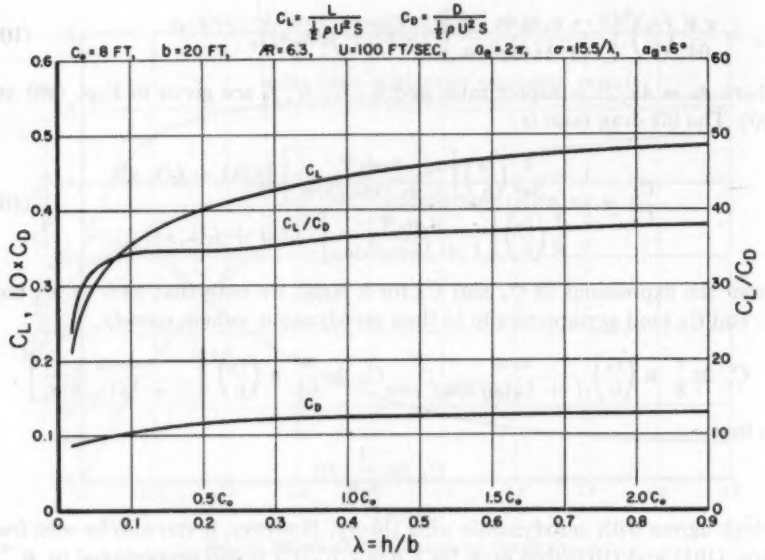

$C_L = \frac{L}{\frac{1}{2}\rho U^2 S}$ $C_D = \frac{D}{\frac{1}{2}\rho U^2 S}$
$c_0$ = 8 FT, b = 20 FT, AR = 6.3, U = 100 FT/SEC, $a_0$ = 2π, σ = 15.5/λ, $\alpha_0$ = 6°
$C_L$
$C_L/C_D$
$C_D$
0.5 $C_0$
1.0 $C_0$
1.5 $C_0$
2.0 $C_0$
$C_L$, 10×$C_D$
$C_L/C_D$
λ = h/b

FIG. 6. Effect of depth on $C_L$, $C_D$ and $C_L/C_D$ (large $\sigma$).

Let us first investigate the case of large $\sigma$. Neglecting the possibility of cavitation, we shall take the operating condition

$$U = 100 \text{ ft/sec}, \qquad \sigma = U^2/gh = 15.5/\lambda. \tag{108}$$

The value of $\sigma$ may be regarded as large for $\lambda < 2$ ($\sigma > 7.7$). Then we may use the formulas of $f_1$, $f_2$, $f_4$ and $f_6$ for $\sigma$ large. The results are plotted against $\lambda$ for $\alpha_0 = 6°$ in Fig. 6 and also plotted against $\alpha_0$ for several small values of $\lambda$ in Fig. 7. From these curves several interesting conclusions may be drawn.

At depths greater than 2 chords, the influence of the surface of the water is negligibly small and the hydrofoil will have characteristics similar to those of an airfoil of the same section. In the range of depths less than 2 chords, $C_D$ decreases gradually with decreasing $\lambda$ because the decrease of liquid flow (with respect to $U$) above the foil diminishes the induced downwash. At very small depths ($h < \frac{1}{2}c_0$), the wave drag only is important; our results show that when the hydrofoil is near the surface, the rate of energy shed to form the wave system is slightly less than the induced drag at large depths. In the same range of operation ($h < 2c_0$), $C_L$ decreases comparatively rapidly to almost $\frac{1}{2}C_{L_\infty}$ as the hydrofoil approaches the surface. This reduction in lift results from the decrease of mass of water flowing over the upper surface of the hydrofoil, causing a reduction of the absolute value of the negative pressure on the suction side. The corresponding lift-drag ratio decreases very slowly with decrease in depths for $h > \frac{1}{2}c_0$ and is almost constant about the point $h = c_0/2$. This ratio decreases rapidly with further decrease in depth for $h < \frac{1}{2}c_0$. This result also indicates

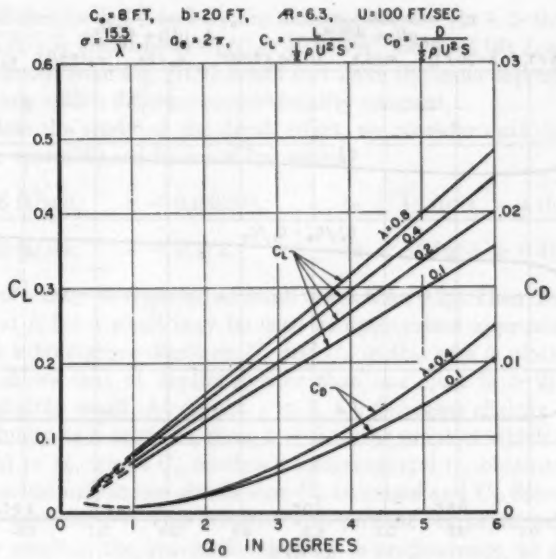



FIG. 7. Hydrodynamic characteristics of the hydrofoil at shallow submergence (large $\sigma$).

that if the water surface is not too choppy, then the depths between $\frac{1}{2}c_0$ and $1c_0$ would be a favorable range for operation. This optimum range of depths also associates with a stabilizing effect because this range corresponds to the middle part of increasing slope $dC_L/d\lambda$; consequently, any further decrease in depths will result in a rapid decrease of lift so the hydrofoil will sink, and if it sinks, an increase in lift will raise it up again. The above results as shown in Fig. 6 agree well with some of previous experiments (cf. Ref. 18, Fig. 15, and Refs. 19, 5).

If we add an almost constant frictional drag (about $C_{D_f} = 0.02$) to the drag coefficient, then the result shown in Fig. 7 is in good agreement with the observations made in Ref. 22 and 19. The reduction in $C_L$ and $C_D$ with decrease in depths becomes more appreciable at larger angles of attack.

It is also of interest to consider the situation of practical operation. When a hydrofoil moves through water of infinite surface extent at shallow submergence, an observer who moves with the hydrofoil can measure the value of the approaching flow velocity $U + u(0)$ more easily (for instance, with a pitot tube) than the value of $U$ at upstream infinity. Then if one defines the over-all lift and drag coefficient based on the value of $U + u(0)$, that is,

$$C'_L = \frac{L}{\frac{1}{2}\rho U^2[1 + u(0)/U]^2 S}, \qquad C'_D = \frac{D}{\frac{1}{2}\rho U^2[1 + u(0)/U]^2 S} \tag{109a}$$

where

$$\frac{u(0)}{U} = -\frac{1}{2\pi}\frac{(c_0/b)a_e\alpha_a}{1 + \frac{1}{8}(c_0/b)a_e}[\Psi f_1(\lambda, \sigma)], \tag{109b}$$

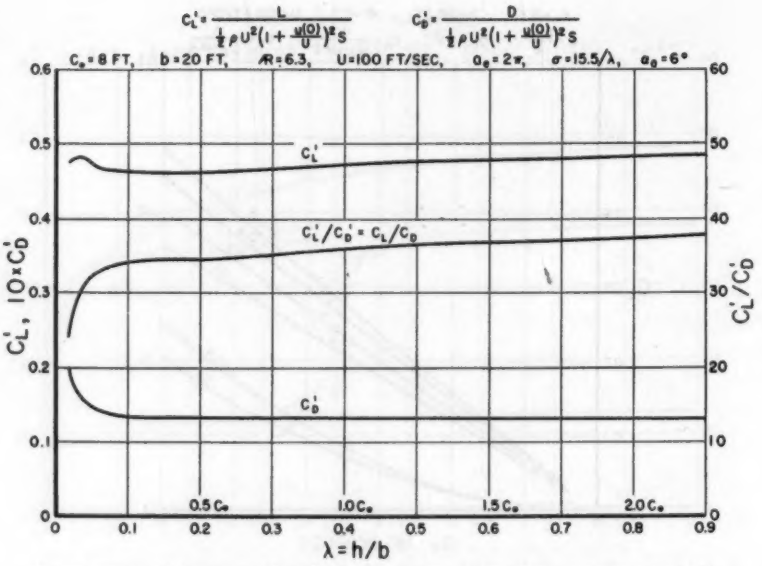

$C_L' = \frac{L}{\frac{1}{2}\rho U^2(1+\frac{u(0)}{U})^2 S}$
$C_D' = \frac{D}{\frac{1}{2}\rho U^2(1+\frac{u(0)}{U})^2 S}$
$C_0$ = 8 FT, b = 20 FT, $\mathcal{R}$ = 6.3, U = 100 FT/SEC, $a_0 = 2\pi$, $\sigma = 15.5/\lambda$, $a_0$ = 6°
$C_L'$
$C_L'/C_D' = C_L/C_D$
$C_D'$
0.5 $C_0$
1.0 $C_0$
1.5 $C_0$
2.0 $C_0$
$C_L'$, 10×$C_D'$
$C_L'/C_D'$
$\lambda = h/b$

FIG. 8. Effect of depth on $C_L'$, $C_D'$ (based on approaching flow velocity, large $\sigma$).

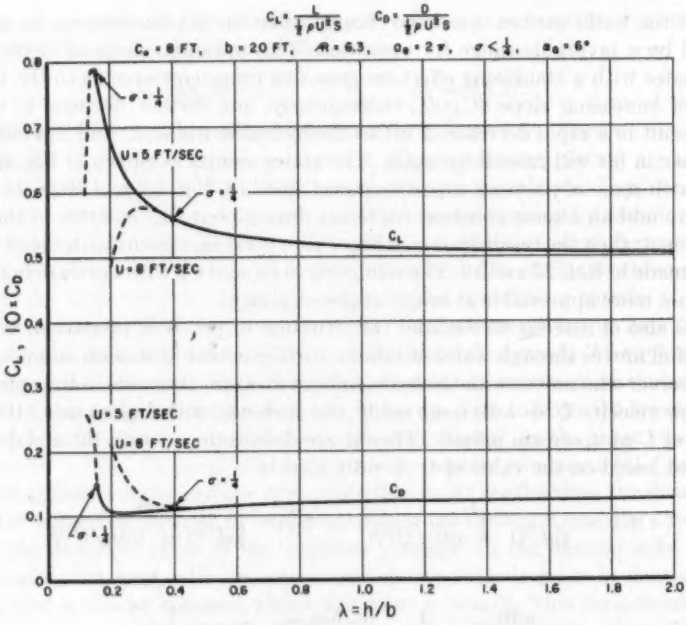

$C_L = \frac{L}{\frac{1}{2}\rho U^2 S}$ $C_D = \frac{D}{\frac{1}{2}\rho U^2 S}$
$C_0$ = 8 FT, b = 20 FT, $\mathcal{R}$ = 6.3, $a_0 = 2\pi$, $\sigma < \frac{1}{4}$, $a_0$ = 6°
$\sigma = \frac{1}{4}$
U = 5 FT/SEC
U = 8 FT/SEC
$C_L$
$C_D$
$C_L$, 10×$C_D$
$\lambda = h/b$

FIG. 9. Effect of depth on $C_L$ and $C_D$ (small $\sigma$).

one will find that both $C'_L$ and $C'_D$ are almost constant for $\lambda > 0.1$ (cf. Fig. 8) although $C'_L/C'_D$ is identical to $C_L/C_D$. However, the total lift $L$ and the total drag $D$ calculated from Eq. (109) would still have the same dependence on $\lambda$ as $C_L$ and $C_D$ only with a different proportionality constant.

To complete the study of the depth effect, we consider next, in the case of small $\sigma$, two operating conditions of low-speed:

(A) $U = 5$ ft/sec, $\sigma = 0.0388/\lambda$, ($\sigma < \frac{1}{4}$ for $\lambda > 0.16$); (110a)

(B) $U = 8$ ft/sec, $\sigma = 0.1/\lambda$, ($\sigma < \frac{1}{4}$ for $\lambda > 0.4$). (110b)

The values of $\sigma$ may be regarded as small if less than 0.25. Then the formulas of $f_1$, $f_2$, $f_4$ and $f_6$ for $\sigma$ small may be used as appropriate approximations. The effect of the submergence depth on $C_L$ and $C_D$ in this case is plotted in Fig. 9. The result shows that at depths greater than one span ($\lambda > 2$) the surface effect is negligibly small. At depths $\lambda < 2$, $C_D$ decreases slightly while $C_L$ increases gradually as $\lambda$ decreases from $\lambda = 2$ to the point at which $\sigma$ is approximately equal to $\frac{1}{4}$, where $C_D$ reaches a minimum and $C_L$ obtains a maximum value. Numerical calculation shows that $C_D$ increases and $C_L$ decreases rapidly with further decrease in $\lambda$. This behavior may probably be explained by visualizing that for small $\sigma$, the gravitational effect is predominant, so that the free

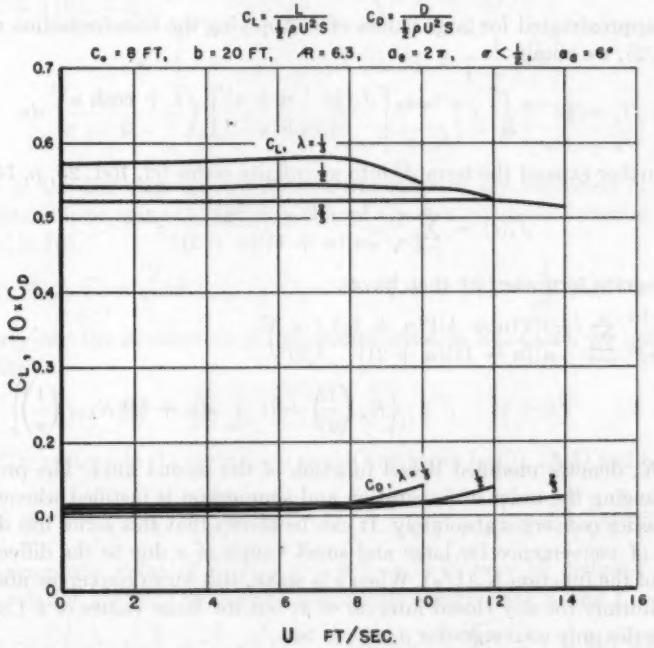

$C_L = \frac{L}{\frac{1}{2}\rho U^2 S}$ $C_D = \frac{D}{\frac{1}{2}\rho U^2 S}$
$C_o$ = 8 FT, b = 20 FT, $\mathcal{R}$ = 6.3, $a_o = 2\pi$, $\sigma < \frac{1}{2}$, $\alpha_g$ = 6°
$C_L$, $\lambda = \frac{1}{2}$
$C_D$, $\lambda = \frac{1}{2}$
$C_L$, 10 × $C_D$
U FT/SEC.
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 2 4 6 8 10 12 14 16

FIG. 10. Effect of free stream velocity $U$ on $C_L$, $C_D$ at small values of $\sigma$.

surface is almost immobile and acts like a solid boundary to which one may draw the analogy of so-called ground effect in aerodynamics.

From the above result derived from this analysis we find that the effect of free stream velocity $U$ on $C_L$ and $C_D$ becomes significant only for motions with small Froude number and small $\lambda$. For instance, if we take three small values of $\lambda$, say, $\lambda = \frac{1}{3}, \frac{1}{2}$ and $\frac{2}{3}$, then $\sigma$ of the aforementioned hydrofoil may be taken as small for $U$ less than 10, 12, and 14 ft/sec respectively. Using Eqs. (105)–(107) together with (93c, e), (98d) and (99d), we obtain the result as shown by Fig. 10. The corresponding values of $C_L$ and $C_D$ are almost constant with respect to $U$ for $\sigma$ approximately less than $\frac{1}{4}$; and for greater values of $\sigma$, $C_L$ decreases and $C_D$ increases slightly with increase in $U$.

If one wants to investigate all different effects for intermediate values of $\sigma$, one has to evaluate integrals in Eqs. (63), (72), (82) and (87d) by numerical methods.

**Appendix.** *Evaluation of the Integral Representing Wave Drag.* In this appendix the value of the integral in Eq. (63)

$$I_1 \equiv \int_0^{\pi/2} e^{-(2/\sigma)\sec^2\theta} \left[\frac{J_1\left(\frac{1}{\beta}\sec^2\theta\sin\theta\right)}{\sec^2\theta\sin\theta}\right]^2 \sec^5\theta\, d\theta \tag{A.1}$$

will be approximated for large values of $\sigma$. Applying the transformation $\sec\theta = \cosh(u/2)$, we obtain

$$I_1 = \tfrac{1}{2}e^{-1/\sigma}\int_0^\infty e^{-\sigma^{-1}\cosh u}\left[\frac{J_1(\frac{1}{2}\beta^{-1}\sinh u)}{\frac{1}{2}\sinh u}\right]^2\left(\frac{1+\cosh u}{2}\right)^2 du$$

If we further expand the term $J_1^2$ into an infinite series (cf. Ref. 25, p. 147)

$$J_1^2(z) = \sum_{n=0}^{\infty}\frac{(-)^n\Gamma(n+\frac{3}{2})}{\sqrt{\pi}n!(n+1)!(n+2)!}z^{2n+2}$$

and integrate termwise, we then have

$$I_1 = \frac{e^{-1/\sigma}}{4\pi\beta^2}\sum_{n=0}^{\infty}\frac{(-)^n\Gamma(n+\frac{1}{2})\Gamma(n+\frac{3}{2})}{n!(n+1)!(n+2)!}\left(\frac{\sigma}{2\beta^2}\right)^n \cdot\left\{K_n\left(\frac{1}{\sigma}\right)+[1+\sigma(n+\tfrac{1}{2})]\,K_{n+1}\left(\frac{1}{\sigma}\right)\right\} \tag{A.2}$$

where $K_n$ denotes modified Bessel function of the second kind. The process of interchanging the order of integration and summation is justified whenever the above series converges absolutely. It can be shown that this series has different regions of convergence for large and small values of $\sigma$ due to the different behavior of the function $K_n(1/\sigma)$. When $\sigma$ is small, this series converges absolutely and uniformly for any closed interval of $\beta$; but for large values of $\sigma$ ($\gg 1$), the above series only converges for $\beta/\sigma = \lambda > 1$.

Now the sum of the series in Eq. (A.2) (also cf. Eq. 64) will be approximated

for large values of $\sigma$. In this case $K_n(1/\sigma)$ has the expansion given by Eq. (67). It is easy to see that the most important contribution to the sum of this series comes from the term $\frac{1}{2}(n!)(2\sigma)^{n+1}$ of the expansiom of $K_{n+1}(1/\sigma)$. Here it seems convenient to decompose $I_1$ into three parts

$$I_1 = I_{11} + I_{12} + I_{13} \tag{A.3}$$

where $I_{11}$, $I_{12}$ and $I_{13}$ are given below.

$$\begin{aligned} I_{11} &\equiv \frac{e^{-1/\sigma}}{4\pi\beta^2} \sum_{n=0}^{\infty} \frac{(-)^n \Gamma(n+\frac{1}{2})\Gamma(n+\frac{3}{2})}{n!(n+1)!(n+2)!} \left(\frac{\sigma}{2\beta^2}\right)^n [\sigma(n+\tfrac{1}{2})]\, K_{n+1}\left(\frac{1}{\sigma}\right) \\ &\cong \frac{e^{-1/\sigma}}{4\pi} \left\{ \sum_{n=0}^{\infty} \frac{(-)^n \Gamma^2(n+\frac{3}{2})}{(n+1)!(n+2)!} \lambda^{-2(n+1)} \right. \\ &\qquad \left. - \frac{1}{4\beta^2} \sum_{n=1}^{\infty} \frac{(-)^n \Gamma^2(n+\frac{3}{2})}{n(n+1)!(n+2)!} \lambda^{-2n} - \frac{\pi \log(2\sigma)}{16\beta^2} \right\} \end{aligned} \tag{A.4}$$

where the expansion formula Eq. (67) has been used. Now both these two series converge only when $\lambda \geqq 1$; but the first series can be directly related to a hypergeometric function which may further be continued analytically to the whole region $0 \leqq \lambda < \infty$ as follows:

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(-)^n \Gamma^2(n+\frac{3}{2})}{(n+1)!(n+2)!} \lambda^{-2(n+1)} &= \pi \left[ 1 - F\left(\tfrac{1}{2}, \tfrac{1}{2}; 2; -\frac{1}{\lambda^2}\right) \right] \\ &= \pi \left[ 1 - \frac{\lambda}{\sqrt{1+\lambda^2}} F\left(\tfrac{1}{2}, \tfrac{3}{2}; 2; \frac{1}{1+\lambda^2}\right) \right] \\ &= \pi \left\{ 1 - \frac{4}{\pi} \lambda \sqrt{1+\lambda^2} \left[ K\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - E\left(\frac{1}{\sqrt{1+\lambda^2}}\right) \right] \right\} \end{aligned} \tag{A.5}$$

where in the last step the hypergeometric function are expressed in terms of complete elliptic integrals by using one of Gauss's recursion formulas (cf. Ref. 26, pp. 9, 10).

$$F(\tfrac{1}{2}, \tfrac{3}{2}; 2; k^2) = \frac{2}{k^2} [F(\tfrac{1}{2}, \tfrac{1}{2}; 1; k^2) - F(-\tfrac{1}{2}, \tfrac{1}{2}; 1; k^2)] = \frac{4}{\pi k^2} [K(k) - E(k)].$$

To carry out the summation of the second series in Eq. (A.4), we note that if we write

$$f(z) = \sum_{n=1}^{\infty} \frac{(-)^n \Gamma^2(n+\frac{3}{2})}{n(n+1)!(n+2)!} z^{-n}, \qquad z = \lambda^2, \tag{A.6}$$

then $f'(z)$ reduces to the same form as the first series in Eq. (A.4) and hence

$$f'(z) = -\pi \left\{ 1 - \frac{1}{8z} - \frac{4}{\pi} \sqrt{z(1+z)} \left[ K\left(\frac{1}{\sqrt{1+z}}\right) - E\left(\frac{1}{\sqrt{1+z}}\right) \right] \right\}.$$

The asymptotic value of $f'(z)$ as $z \to 0$ ($\sim\lambda \to 0$) may be deduced from the above relation by using the known asymptotic expansion of $K$ and $E$

$$f'(z) \cong \pi \left\{ -1 + \frac{1}{8z} + \frac{4}{\pi} \sqrt{\frac{1}{1+z}} \left[ \log 4 \sqrt{\frac{1+z}{z}} - 1 + 0(z \log z) \right] \right\}.$$

Since the integral of an asymptotic expression is also asymptotic, the asymptotic value of $f(z)$ is then

$$f(z) \cong \frac{\pi}{8} \log z + 0(1) = \frac{\pi}{4} \log \frac{\beta}{\sigma} + 0(1), \quad \text{as } \frac{1}{\sigma} \to 0. \tag{A.7}$$

Therefore, substituting Eqs. (A.5, 7) into (A.4), we obtain

$$I_{11} \cong \frac{1}{4}\left\{1 - \frac{4}{\pi}\lambda\sqrt{1+\lambda^2}\left[K\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - E\left(\frac{1}{\sqrt{1+\lambda^2}}\right)\right] - \frac{1}{16\beta^2}\log\beta + 0\left(\frac{1}{\beta^2}, \frac{1}{\sigma}\right)\right\} \tag{A.8}$$

The second part of $I_1$, namely, $I_{12}$, is given by

$$\begin{aligned} I_{12} &\equiv \frac{e^{-1/\sigma}}{4\pi\beta^2} \sum_{n=0}^{\infty} \frac{(-)^n \Gamma(n+\frac{1}{2})\Gamma(n+\frac{3}{2})}{n!(n+1)!(n+2)!} \left(\frac{\sigma}{2\beta^2}\right)^n K_{n+1}\left(\frac{1}{\sigma}\right) \\ &\cong \frac{e^{-1/\sigma}}{4\pi\sigma} \sum_{n=0}^{\infty} \frac{(-)^n \Gamma(n+\frac{1}{2})\Gamma(n+\frac{3}{2})}{(n+1)!(n+2)!} \left(\frac{\sigma}{\beta}\right)^{2(n+1)} \\ &\cong \frac{e^{-1/\sigma}}{2\sigma}\left[F\left(-\tfrac{1}{2}, \tfrac{1}{2}; 2; -\frac{1}{\lambda^2}\right) - 1\right] \\ &\cong \frac{e^{-1/\sigma}}{2\beta}\left[\sqrt{1+\lambda^2}\, F\left(-\tfrac{1}{2}; \tfrac{3}{2}; 2; \frac{1}{1+\lambda^2}\right) - \lambda\right] \\ &\cong \frac{1}{3\beta}\left\{\frac{2}{\pi}(1+\lambda^2)^{3/2} E\left(\frac{1}{\sqrt{1+\lambda^2}}\right) \right. \\ &\qquad \left. - \tfrac{3}{2}\lambda - \lambda^2\sqrt{1+\lambda^2}\, F\left(-\tfrac{1}{2}, \tfrac{3}{2}; 1; \frac{1}{1+\lambda^2}\right)\right\}\left(1 + 0\left(\frac{1}{\sigma}\right)\right) \end{aligned} \tag{A.9}$$

where in the last step, after the analytic continuation, Gauss's recursion formula is again used. It is evident that the last hypergeometric function in Eq. (A.9) has singularity at $\lambda = 0$. To calculate its value near $\lambda = 0$, we need the following formula

$$F(a, b; a+b; z) = \sum_{n=0}^{\infty} \alpha_n (1-z)^n [\beta_n + \log(1-z)]$$

where

$$\alpha_n = -\frac{\Gamma(a+b)\Gamma(a+n)\Gamma(b+n)}{\Gamma^2(a)\Gamma^2(b)\Gamma^2(n+1)},$$

$$\beta_n = \psi(a+n) + \psi(b+n) - 2\psi(1+n).$$

Hence

$$F\left(-\tfrac{1}{2}, \tfrac{3}{2}; 1; \frac{1}{1+\lambda^2}\right) = \frac{2}{\pi}\left(2 - \log\frac{4\sqrt{1+\lambda^2}}{\lambda}\right) + 0(\lambda^2 \log\lambda)$$

and finally we have

$$I_{12} \cong \frac{2}{3\pi\beta}\left\{(1+\lambda^2)^{3/2} - \frac{3\pi}{4}\lambda - \tfrac{3}{2}\lambda^2\sqrt{1+\lambda^2}\log\frac{4\sqrt{1+\lambda^2}}{\lambda} + 0(\lambda^2)\right\}. \tag{A.10}$$

The third part of $I_1$ denoted by $I_{13}$ is

$$I_{13} \equiv \frac{e^{-1/\sigma}}{4\pi\beta^2} \sum_{n=0}^{\infty} \frac{(-)^n \Gamma(n+\frac{1}{2})\Gamma(n+\frac{3}{2})}{n!(n+1)!(n+2)!} \left(\frac{\sigma}{2\beta^2}\right)^n K_n\left(\frac{1}{\sigma}\right)$$

$$\cong \frac{e^{-1/\sigma}}{8\pi\beta^2} \left\{ \sum_{n=0}^{\infty} \frac{(-)^n \Gamma(n+\frac{1}{2})\Gamma(n+\frac{3}{2})}{n(n+1)!(n+2)!} \lambda^{-2n} + \frac{\pi}{2} \log 2\sigma \right\}.$$

The series in the above equation can be approximated in a manner similar to that applied to Eq. (A.6). The final result is

$$I_{13} \cong \frac{1}{16\beta^2} \log \beta + 0\left(\frac{1}{\beta^2}, \frac{1}{\sigma}\right).$$

Therefore, substituting Eqs. (A.8, 10, 11) into Eq. (A.3), we obtain a good approximation of the value of $I_1$ for $\sigma$ large as follows:

$$I_1 \cong \frac{1}{4}\left\{1 - \frac{4}{\pi}\lambda\sqrt{1+\lambda^2}\left[K\left(\frac{1}{\sqrt{1+\lambda^2}}\right) - E\left(\frac{1}{\sqrt{1+\lambda^2}}\right)\right]\right\}$$

$$+ \frac{1}{3\beta}\left\{\frac{2}{\pi}(1+\lambda^2)^{3/2} E\left(\frac{1}{\sqrt{1+\lambda^2}}\right)\right. \tag{A.12}$$

$$\left. - \tfrac{3}{2}\lambda - \lambda^2\sqrt{1+\lambda^2}\, F\left(-\tfrac{1}{2}, \tfrac{3}{2}; 1; \frac{1}{1+\lambda^2}\right)\right\} + \frac{3}{64\beta^2}\log\beta + 0\left(\frac{1}{\beta^2}, \frac{1}{\sigma}\right).$$

## REFERENCES


1. Y. T. Wu, "A Theory for Hydrofoils of Finite Span", Hydrodynamics Laboratory Report No. 26-8. California Institute of Technology, California (1953).
2. A. Guidoni, "Seaplanes—Fifteen Years of National Aviation", Jour. R. A. S., Vol. 32, No. 205, 25-64, (1928).
3. M. W. Keldysch and M. A. Lawrentjew, "On the Motion of a Wing below the Surface of a Heavy Fluid", ZAHI paper, (Jan. 1935) Moscow.
4. N. E. Kotschin, "On the Motion of Profiles of Any Form below the Surface of a Heavy Fluid", ZAHI paper, (Dec. 1935) Moscow.
5. A. Wladimirow, "Approximative Hydrodynamic Calculation of a Hydrofoil of Finite Span," ZAHI Rep. No. 311, (1937) Moscow (English translation available, Central Air Documents Office No. ATI-57689).
6. H. Lamb, "Hydrodynamics", 6th ed., p. 410, Dover publication New York.
7. T. H. Havelock, "Some Cases of Wave Motion due to a Submerged Obstacle", Proc. Roy. Soc. (A), Vol. 93, 520-532 (1917).
8. T. H. Havelock, "Wave Resistance, Some Cases of Three-dimensional Fluid Motion", Pro. Roy. Soc., Vol. 95, 354-365 (1918).
9. J. Ackeret, "Kavitation and Kavitationskorrosion Hydromechanische Probleme", Hamb. S. V. Werft Reederei Hafen, Vol. 13 n. 16/17, 239-240, 253-255 (1932), Also NACA TM No. 1078 (1945).
10. O. Walchner, "Profilmessungen bei Kavitation Hydromechanische Probleme", Hamb. S. V. Werft Reederei Hafen, Vol. 13 n. 17, 251-252 (1932), Also NACA TM No. 1060 (1944).
11. H. Lerbs, "Untersuchung der Kavitation an Schraubenpropellern", Mitteilung der Hamb. S. V. Werft Reederei Hafen, (1931).
12. E. Martyrer, "Kraftmessungen an Fluegelprofilen bei Kavitation Hydromechanische Probleme", Hamb. S. V. (1932).

13. F. Gutsche, "Versuche ueber die Profileigenschaften der Blattschnitte von Schiffsschrauben", Mitteilung der Versuch. f. Wasserbau u. Schiffbau, 10, (1933).
14. L. P. Smith, "Cavitation on Marine Propellers", Trans. A.S.M.E. Vol. 59, 5, 409–431 (1937).
15. F. Weinig, "Zur Theorie des Unterwassertragflügels und der Gleitfläche", Luftfahrtforschung Vol. 14, 6, 314–324 (1937). Also NACA TM No. 845 (1938).
16. O. Tietjens, "Das Tragflaechenboot", J. d. Luftf. Forschg. Bd. Vol. 1, 361–370 (1937); Werft Reederei Hafen Jahrg. 18, Heft 7, 87–90 (1937); Heft 8, 106–109 (1937).
17. von Schertel, "Versuche mit Tragflaechen-Schnellbooten", Tagungsbericht 066/003 der Lilienthal-Ges.
18. K. E. Ward and N. S. Land, "Preliminary Tests in the NACA Tank to Investigate the Fundamental Characteristics of Hydrofoils", NACA Wartime Rept. L-766 (1940).
19. J. M. Benson and N. S. Land, "An Investigation of Hydrofoils in the NACA Tank", NACA Wartime Rept. L-758 (1942).
20. N. S. Land, "Characteristics of an NACA 66, S-209 Section Hydrofoil at Several Depths", NACA Wartime Rept. L-757 (1943).
21. F. H. Imlay, "Theoretical Motions of Hydrofoil Systems", NACA TN No. 1285 (1947).
22. W. Sottorf, "Experimentelle Untersuchungen zur Frage des Wassertragflügels", Deutsche Versuchsanstalt für Luftfahrt E. V., Hamburg, Rep. No. 408/1 (1940) (English translation available, Central Air Document Office No. ATI 64631, 1950).
23. Th. von Kármán, "Neue Darstellung der Tragflügeltheorie", ZAMM, Nr. 1/2, 56–61 (1935).
24. H. Jeffreys and B. S. Jeffreys, "Methods of Mathematical Physics", 2nd ed., Cambridge University Press (1950).
25. G. N. Watson, "Bessel Functions", Cambridge University Press, London (1948).
26. W. Magnus and F. Oberhettinger, "Special Functions of Mathematical Physics", Chelsea Publishing Co. (1949).
27. E. Jahnke and F. Emde, "Tables of Functions", 4th ed. Dover Publication.
28. I. Lotz, Zeitschrift f. Flugtechnik u. Motorluftschiff. 22, p. 189 (1931).
29. H. Glauert, "The Elements of Airfoil and Airscrew Theory", Cambridge University Press, p. 138 (1930).
30. J. S. Ausman, "Pressure Limitation on the Upper Surface of a Hydrofoil", Ph.D. Thesis, University of California, Berkeley, Calif. (1953).

California Institute of Technology
(Received November 10, 1953)

# A NOTE ON THE ASYMPTOTIC SOLUTIONS OF THE EQUATION OF HYDRODYNAMIC STABILITY*

RICHARD C. DI PRIMA†

**1. Introduction.** For an incompressible, viscous, essentially parallel flow the linearized differential equation for the amplitude function of two dimensional disturbances is, in dimensionless form,

$$\phi^{IV} - 2\alpha^2\phi^{II} + \alpha^4\phi - i\alpha R\{(w - c)(\phi^{II} - \alpha^2\phi) - w''\phi\} = 0. \tag{1.1}$$

Here, the disturbance is given by the stream function $\psi = \phi(y) \exp\{i\alpha(x - ct)\}$, $w(y)$ is the basic velocity distribution, and $R$ is the Reynolds number of the basic flow.

To solve the associated characteristic value problems, asymptotic representations of a fundamental system of four solutions $\phi_I$, $\phi_{II}$, $\phi_{III}$, $\phi_{IV}$ of Eq. (1.1) are needed. Two solutions $\phi_1$ and $\phi_2$ are obtained from the inviscid equation.

$$(w - c)(\phi^{II} - \alpha^2\phi) - w''\phi = 0. \tag{1.2}$$

Tollmien [1] has applied the usual theory of differential equations to study solutions of Eq. (1.2) in the neighborhood of the critical point $y_0$, $w(y_0) = c$, obtaining

$$\phi_1(y) = yB_1(y), \tag{1.3}$$

$$\phi_2(y) = B_2(y) + (w_0''/w_0')yB_1(y) \log y, \tag{1.4}$$

where for simplicity $y_0$ is taken to be zero. The power series for $B_1(y)$ and $B_2(y)$ are

$$B_1(y) = 1 + (w_0''/2w_0')y + (w_0'' + \alpha^2)y^2 + \cdots, \tag{1.5}$$

$$B_2(y) = 1 + [-(w_0''/w_0'^2) + (w_0'''/2w_0') + \tfrac{1}{2}\alpha^2]y^2 + \cdots. \tag{1.6}$$

The $\phi_1$ solution is regular, but the $\phi_2$ solution has a logarithmic singularity with derivative becoming infinite as $y \to 0$, and hence its derivatives fail to represent asymptotically those of the solution $\phi_{II}$ of Eq. (1.1). In a boundary value problem it is actually necessary to consider the case in which $y \to 0$ as the small parameter $\epsilon = (\alpha R)^{-1/3} \to 0$, more precisely $\eta = y/\epsilon$ approaches a finite non-zero value as $\epsilon \to 0$. Recently Tollmien (considering only real $y$ and real parameters $\alpha$, $R$, and $c$) constructed an improved solution $\phi_2^*$ which is an approximation to $\phi_{II}$ even in the neighborhood of zero [2]. For $y = 0(\epsilon)$, the most difficult case, he states the following error estimates

$$|\phi_{II} - \phi_2^*| = 0(\epsilon),$$

$$|d\phi_{II}/dy - d\phi_2^*/dy| = 0(\epsilon \log \epsilon).$$

* This work was sponsored by the Office of Naval Research under Contract N 5 ori-07871.

† The author would like to express his appreciation to Professor C. C. Lin for suggesting this work, and for his continued interest during many helpful discussions. He is also indebted to Dr. Wasow for helpful comments for the final version.

Wasow (considering general values of $y$ and the parameters) has given a complete theory of the manner in which the asymptotic solutions represent the fundamental solutions of Eq. (1.1) [3]. However, his error estimate for the derivative of the solution $v_3$ which approximates $\phi_{II}$ is not as good as that stated by Tollmien for $d\phi_2^*/dy$.

In this paper the representations of $\phi_2^*$, $v_3$ and their derivatives are studied in greater detail than previously for the case when $\eta$ is finite, and when $|\eta| \to \infty$. Using these representations, it will be shown that the agreement between $\phi_2^*$ and $v_3$ and their first derivatives is $0(\epsilon)$ as long as $\eta$ does not equal zero, i.e., for the range of $\eta$ of interest in the eigenvalue problem. Thus one suspects that $v_3$ and its derivative approximate $\phi_{II}$ and $d\phi_{II}/dy$ to the same degree of accuracy for complex $c$ as does the Tollmien solution for real $c$.

**2. Tollmien Solution.** Tollmien [2] considers in place of $\phi_2$ the function $\phi_2^*$ defined by

$$\phi_2^* = B_2(y) + B_1(y)\epsilon_T\phi_{21}(\eta_T), \tag{2.1}$$

where $\eta_T = y(\alpha R w_0')^{1/3} = y/\epsilon_T$, and $\phi_{21}(\eta_T)$ satisfies the following differential equation:

$$-i\frac{d^4\phi_{21}}{d\eta_T^4} = \eta_T\frac{d^2\phi_{21}}{d\eta_T^2} - \frac{w_0''}{w_0'}, \tag{2.2}$$

and certain conditions to be specified. If Eq. (2.2) is integrated for $\phi_{21}''(\eta_T)$, there are two constants of integration. These are determined so that $\phi_{21}''(\eta_T) \to 0$ as $\eta_T \to \pm\infty$. Then it is found that

$$\frac{d^2\phi_{21}}{d\eta_T^2} = \frac{\pi}{6}\frac{w_0''}{w_0'}\eta_T^{\frac{1}{2}}\left\{H_{\frac{1}{3}}^{(2)}\int_\infty^{\eta_T}\eta_T^{\frac{1}{2}}H_{\frac{1}{3}}^{(1)}\,d\eta_T - H_{\frac{1}{3}}^{(1)}\int_{-\infty}^{\eta_T}\eta_T^{\frac{1}{2}}H_{\frac{1}{3}}^{(2)}\,d\eta_T\right\}, \tag{2.3}$$

where the argument of the Hankel functions is $(\frac{2}{3})(i\eta_T)^{3/2}$. In obtaining $\phi_{21}(\eta_T)$ from Eq. (2.3) there are two more constants of integration to be determined.

Before discussing the determination of these constants it is desirable to have the asymptotic representation of $d^2\phi_{21}/d\eta_T^2$. Making the change of variables $d^2\phi_{21}/d\eta_T^2 = \eta_T^{\frac{1}{2}}\psi(\xi)$, $\xi = (\frac{2}{3})(i\eta_T)^{3/2}$ and using the asymptotic behavior of the Lommel function ([5], p. 347 and 351) it is not difficult to show that for $-2\pi/3 < \arg(i\eta_T) < 2\pi/3$

$$\frac{d^2\phi_{21}}{d\eta_T^2} \sim \frac{w_0''}{w_0'}\left\{\frac{1}{\eta_T} - \sum_{m=1}^{\infty}\frac{(-i)^m(-3m)(3m-1)!}{3^m\cdot m!}\eta_T^{-3m-1}\right\}. \tag{2.4}$$

Now we may write

$$\phi_{21}(\eta_T) = \bar{\phi}_{21}(\eta_T) + A\eta_T + B, \tag{2.5}$$

where $\bar{\phi}_{21}(\eta_T)$ is an analytic function for all finite $\eta_T$ defined by

$$\begin{aligned}\frac{d\bar{\phi}_{21}}{d\eta_T} &= \int_\infty^{\eta_T}\left(\frac{d^2\phi_{21}}{d\eta_T^2} - \frac{w_0''}{w_0'}\frac{1}{\eta_T}\right)d\eta_T + \frac{w_0''}{w_0'}\log\eta_T,\\ \bar{\phi}_{21} &= \int_\infty^{\eta_T}\left(\frac{d\bar{\phi}_{21}}{d\eta_T} - \frac{w_0''}{w_0'}\log\eta_T\right)d\eta_T + \frac{w_0''}{w_0'}(\eta_T\log\eta_T - \eta_T).\end{aligned} \tag{2.6}$$

The constants $A$ and $B$ are determined by requiring that $\phi_2^* - \phi_2$ and $d(\phi_2^* - \phi_2)/dy$ are of $0(\epsilon_T)$ for finite $y$. From Eqs. (1.3), (1.4), (2.1) and (2.5) we have

$$\phi_2^* - \phi_2 = \epsilon_T B_1(y)\{\bar{\phi}_{21}(\eta_T) - (w_0''/w_0')\eta_T \log \eta_T\} + \{A - (w_0''/w_0') \log \epsilon_T\} y B_1(y) + \epsilon_T B \cdot B_1(y),$$

and

$$\frac{d}{dy}(\phi_2^* - \phi_2) = B_1(y)\left\{\frac{d\bar{\phi}_{21}}{d\eta_T} + A - \frac{w_0''}{w_0'}(1 + \log y)\right\} + \epsilon_T\left\{\bar{\phi}_{21}(\eta_T) + A\eta_T - \frac{w_0''}{w_0'}\eta_T \log y + B\right\}. \tag{2.7}$$

It is clear from Eq. (2.4) that the integrals in Eqs. (2.6) behave as negative powers of $\eta_T$ for large $\eta_T$, hence we have the desired agreement for finite $y$ if we take $A = (w_0''/w_0')(1 + \log \epsilon_T)$ and any finite value for $B$.

We shall now show that, with such a choice, the differences $\phi_2^* - \phi_2$ and $d(\phi_2^* - \phi_2)/dy$ are given by

$$\phi_2^* - \phi_2 = 0(\epsilon_T),$$
$$\frac{d}{dy}(\phi_2^* - \phi_2) \sim -\frac{w_0''}{w_0'} B_1(y) \sum_{m=1}^{\infty} \frac{\Gamma(3m)(-i)^m}{3^m \cdot m!} \eta_T^{-3m} + 0(\epsilon_T), \tag{2.8}$$

for large $\eta_T$, $-2\pi/3 < \arg(i\eta_T) < 2\pi/3$, and for finite $\eta_T$ by

$$\phi_2^* - \phi_2 = 0(\epsilon_T),$$
$$\frac{d(\phi_2^* - \phi_2)}{dy} = -\frac{w_0''}{w_0'} B_1(y)\left\{C + \log \eta_T + \sum_{m=1}^{\infty} \frac{3^{\frac{1}{3}m-1}\Gamma(\frac{1}{3}m)}{m!}(-i\eta_T)^m\right\} + 0(\epsilon_T), \tag{2.9}$$

where $C$ is a constant which will be defined later. These relations will be useful for comparison with Wasow's solution, since his solution is related to the inviscid solution $\phi_2$.

Briefly, these results are obtained in the following way. First, the regularity of $\bar{\phi}_{21}$ is used in showing the relations for $\phi_2^* - \phi_2$ to be true for all $\eta_T$. For $d(\phi_2^* - \phi_2)/dy$ it is necessary to have the behavior of the function $d\bar{\phi}_{21}/d\eta_T - (w_0''/w_0') \log \eta_T$, since from the second of Eqs. (2.7)

$$\frac{d(\phi_2^* - \phi_2)}{dy} = B_1(y)\left\{\frac{d\bar{\phi}_{21}}{d\eta_T} - \frac{w_0''}{w_0'} \log \eta_T\right\} + 0(\epsilon_T) \tag{2.10}$$

For large $\eta_T$, $-2\pi/3 < \arg(i\eta_T) < 2\pi/3$ it follows from Eqs. (2.4) and (2.6) that

$$\frac{d\bar{\phi}_{21}}{d\eta_T} - \frac{w_0''}{w_0'} \log \eta_T \sim -\frac{w_0''}{w_0'} \sum_{m=1}^{\infty} \frac{\Gamma(3m)(-i)^m}{3^m \cdot m!} \eta_T^{-3m}. \tag{2.11}$$

To obtain a power series representation for this function for finite $\eta_T$ we write Eq. (2.3) in the following form:

$$\frac{d^2\phi_{21}}{d\eta_T^2} = \frac{\pi}{6}\frac{w_0''}{w_0'}\eta_T^{\frac{1}{2}}\{A_1 H_{\frac{1}{3}}^{(2)} - A_2 H_{\frac{1}{3}}^{(1)}\} + \frac{\pi}{6}\frac{w_0''}{w_0'}\eta_T^{\frac{1}{2}}\left\{H_{\frac{1}{3}}^{(2)}\int_0^{\eta_T}\eta_T^{\frac{1}{2}}H_{\frac{1}{3}}^{(1)}\,d\eta_T - H_{\frac{1}{3}}^{(1)}\int_0^{\eta_T}\eta_T^{\frac{1}{2}}H_{\frac{1}{3}}^{(2)}\,d\eta_T\right\}, \tag{2.12}$$

where

$$A_1 = \int_\infty^0 \eta_T^{\frac{1}{2}}H_{\frac{1}{3}}^{(1)}\,d\eta_T, \qquad A_2 = \int_\infty^0 \eta_T^{\frac{1}{2}}H_{\frac{1}{3}}^{(2)}\,d\eta_T.$$

The second term of Eq. (2.12) can be evaluated by relating it to the Lommel function ([5], p. 346); the first by direct expansion of the Hankel functions; and $A_1$ and $A_2$ by using ([5], p. 39). Carrying out these lengthy, but straightforward calculations we obtain upon integrating from $\infty$ to 1 and then from 1 to $\eta_T$ that

$$\frac{d\bar{\phi}_{21}}{d\eta_T} - \frac{w_0''}{w_0'}\log\eta_T = -\frac{w_0''}{w_0'}\left\{C + \log\eta_T + \sum_{m=1}^{\infty}\frac{i^m 3^{m-1}\Gamma(m)}{(3m)!}\eta_T^{3m} + i3^{-2/3}\sum_{m=0}^{\infty}\frac{i^m 3^m\Gamma(m+\frac{1}{3})}{(3m+1)!}\eta_T^{3m+1} + 3^{-1/3}\sum_{m=0}^{\infty}\frac{i^m 3^m\Gamma(m+\frac{2}{3})}{(3m+2)!}\eta_T^{3m+2}\right\},$$

where the constant $C$ is

$$C = \int_\infty^1\left(\frac{d^2\phi_{21}}{d\eta_T^2} - \frac{w_0''}{w_0'}\frac{1}{\eta_T}\right)d\eta_T - \int_0^1\frac{d^2\phi_{21}}{d\eta_T^2}\,d\eta_T.$$

If for the moment we choose $\arg i = 9\pi/2$ we can write these three power series as one, and for finite $\eta_T$

$$\frac{d\bar{\phi}_{21}}{d\eta_T} - \frac{w_0''}{w_0'}\log\eta_T = -\frac{w_0''}{w_0'}\left\{C + \log\eta_T + \sum_{m=1}^{\infty}\frac{3^{\frac{1}{3}m-1}\Gamma(m/3)}{m!}(-i\eta_T)^m\right\}. \tag{2.13}$$

Since only integral powers of $i$ appear in Eq. (2.13) we may relax our restriction on $\arg i$.

**3. Wasow Solution.** Wasow treats the differential equation

$$u^{(4)}(x) + \sum_{j=1}^{4}a_j(x)u^{(4-j)} + \lambda^2\sum_{j=0}^{2}b_j(x)u^{(2-j)} = 0, \tag{3.1}$$

where $x$ is a complex variable, and $b_0(x)$ has a simple zero taken to be at $x = 0$, also $b_1(x)$ is assumed $\equiv 0$. This equation may be identified with Eq. (1.1) by setting $y - y_0 = kx$, where $w(y_0) = c$ and $k = -w_0''/w_0'$. Then $a_1 \equiv 0$, $a_2 = -2\alpha^2(w_0'/w_0'')^2$, $a_3 \equiv 0$, $a_4 = \alpha^4(w_0'/w_0'')^4$, $b_0 = -(w-c)(w_0''/w_0')^2$, $b_1 \equiv 0$, $b_2 = [(w-c)\alpha^2 + w'']/w_0''$, and $\lambda^2 = -i\alpha R w_0' k^3$. We note that $\lambda^{-2/3} = 0(\epsilon) = 0(\epsilon_T)$. For the sake of convenience in discussing the Wasow solution his notation will be employed. The comparison with the Tollmien solution will be made at the end of this section.

In order to solve Eq. (3.1) Wasow considers the special differential equation

$$\frac{d^4q}{dx^4} + \lambda^2\left(x\frac{d^2q}{dx^2} + q\right) = 0. \tag{3.2}$$

One solution of this equation, valid for $x$ in a sector bounded by $4\pi/3 - \frac{2}{3}\arg\lambda < \arg x < 8\pi/3 - \frac{2}{3}\arg\lambda$, is

$$U_3(x, y) = \int_{C(U_3)} t^{-2}\exp\left\{tx - \frac{1}{t} + \frac{t^3}{3\lambda^2}\right\}dt,$$

where the contour $C(U_3)$ issues from the origin into the right half of the $t$ plane where $Re(t^{-1}) > 0$ and extends to $\infty$ along a ray lying in the sector $5\pi/6 + \frac{2}{3}\arg\lambda < \arg t < 7\pi/6 + \frac{2}{3}\arg\lambda$.

The function $v_3(x, \lambda)$, which approximates $\phi_{II}$, is defined by

$$v_3(x, \lambda) = \frac{{}_0u_3(x)}{{}_0q_3(x)}\,U_3(x, \lambda), \tag{3.3}$$

([3], Eqs. (1.9), (3.5c), (3.16)). Here ${}_0u_3(x)$ is the logarithmic solution of the reduced equation $b_1(x)u'' + b_2(x)u = 0$, (for Eq. (1.1), ${}_0u_3 \equiv \phi_2$). Since $v_3(x, \lambda)$ should behave as ${}_0u_3(x)$, i.e. the inviscid solution, ${}_0q_3(x)$ is specified to be the logarithmic solution of the reduced special equation $xq'' + q = 0$ such that $U_3(x, \lambda)/{}_0q_3(x) = 1 + 0(\epsilon)$. It has been shown by Wasow ([4], Eq. (5.5)) that

$$U_3(x, \lambda) = \pi i\sqrt{x}\,H_1^{(1)}(2\sqrt{x}) + \lambda^{-2/3}E(x, \lambda), \tag{3.4}$$

for all $x^*$ where $E(x, \lambda)$ is a bounded function (the error estimate can be improved for finite $x$). Consequently ${}_0q_3$ is chosen to be $\pi i\sqrt{x}\,H_1^{(1)}(2\sqrt{x})$. See [3] for a more complete description of ${}_0u_3$ and ${}_0q_3$.

From Eq. (3.3) upon using Eq. (3.4) we have for all $x$

$$v_3(x, \lambda) = {}_0u_3(x) + \lambda^{-2/3}E(x, \lambda), \tag{3.5}$$

and upon comparing Eq. (3.5) with Eqs. (2.8) and (2.9) we see that $v_3$ and $\phi_2^*$ agree to terms of $0(\epsilon)$.

It now remains to show that $dv_3/dx$ and $d\phi_2^*/dy$ agree to terms of $0(\epsilon)$ for $\eta_w$ bounded away from zero. It follows from Eq. (3.3), using Eq. (3.4) that

$$\frac{dv_3}{dx} = \frac{d_0u_3}{dx} - \frac{{}_0u_3}{{}_0q_3}\frac{d}{dx}({}_0q_3) + \frac{{}_0u_3}{{}_0q_3}\frac{d}{dx}U_3(x, \lambda) + R(x, \lambda),$$

where

$$R(x, \lambda) = \left(\frac{d_0u_3}{dx} - \frac{{}_0u_3}{{}_0q_3}\frac{d_0q_3}{dx}\right)0(\lambda^{-2/3}).$$

* We are of course thinking of $x$ in the sector defined above. Later in this section the correspondence of regions of validity for the Tollmien and Wasow work will be made. If $w_0'$ is positive then $\arg y = \arg\eta_T$ and $\arg\lambda$ may be determined by requiring that the $x$ and $y$ sectors agree.

To see that $R(x, \lambda)$ is $0(\lambda^{-2/3})$ one writes

$$\frac{d_0 u_3}{dx} - \frac{{}_0u_3}{{}_0q_3}\frac{d_0 q_3}{dx} = {}_0q_3 \frac{d}{dx}\left(\frac{{}_0u_3}{{}_0q_3}\right) = 0(1)$$

because $d/dx({}_0u_3/{}_0q_3)$ is bounded at $x = 0$ (see [3] p. 232 formula (3.16)). From the definition of $U_3(x, \lambda)$ and the fact that

$${}_0q_3(x) = \pi i \sqrt{x}\, H_1^{(1)}(2\sqrt{x}) = \int_{C(U_3)} t^{-2} \exp\{tx - t^{-1}\}\, dt,$$

we deduce that

$$\frac{dv_3}{dx} = \frac{d_0 u_3}{dx} + \int_{C(U_3)} \frac{1}{t} e^{tx-(1/t)}(e^{t^3/3\lambda^2} - 1)\, dt + 0(\lambda^{-2/3}). \tag{3.6}$$

To investigate the integral which appears in Eq. (3.6) we make the change of variables $\sigma^3 = \lambda^{-2}t^3$, and $\eta_w = \lambda^{+2/3}x$, then

$$\int_{C(U_3)} \frac{1}{t} e^{tx-(1/t)}(e^{t^3/3\lambda^2} - 1)\, dt = \int_{C'(U_3)} \frac{1}{\sigma} e^{\sigma\eta_w}(e^{\frac{1}{3}\sigma^3} - 1)\, d\sigma$$
$$+ \int_{C(U_3)} \frac{1}{t} (e^{t^3/3\lambda^2} - 1)e^{tx}(e^{-(1/t)} - 1)\, dt.$$

The first integral on the right hand side is defined for $\eta_w$ not equal to zero and for a path extending from the origin to infinity such that $Re(\sigma^3) < 0$ and $Re(\sigma\eta_w) < 0$. The restrictions on $x$ and $t$ require $5\pi/6 < \arg\ \sigma < 7\pi/6$ and $4\pi/3 < \arg\ \eta_w < 8\pi/3$. The above inequalities may always be satisfied under these conditions. Following an argument analogous to Wasow ([3], p. 229) it can be shown that the second integral on the right hand side is $0(\lambda^{-2/3}\eta_w^{-1})$ and hence is within the order of error in Eq. (3.6). Let

$$H_3(\eta_w) = \int_{C'(U_3)} \frac{1}{\sigma} e^{\sigma\eta_w}(e^{\frac{1}{3}\sigma^3} - 1)\, d\sigma, \tag{3.7}$$

then

$$\frac{dv_3}{dx} = \frac{d_0 u_3}{dx} + \frac{{}_0u_3}{{}_0q_3} H_3(\eta_w) + 0(\lambda^{-2/3}). \tag{3.8}$$

Comparing Eqs. (3.8) and (2.10) we have

$$\frac{d(v_3 - {}_0u_3)}{dx} - \frac{d(\phi_2 - \phi_2^*)}{dy} = \frac{{}_0u_3}{{}_0q_3} H_3(\eta_w) - B_1(y)\left\{\frac{d\bar{\phi}_{21}}{d\eta_T} - \frac{w_0''}{w_0'}\log \eta_T\right\} + 0(\lambda^{-2/3}).$$

To show the desired agreement it is necessary to show that the first two terms on the right hand side of the above equation add up to an expression of order $\lambda^{-2/3}$ for both finite and large $\eta$. Because of the unevaluated constants in Eq. (2.13) this can most conveniently be done by showing that

$$H_3(\eta_w) = \frac{d\bar{\phi}_{21}}{d\eta_T} - \frac{w_0''}{w_0'}\log \eta_T,$$

in which case we obtain Eqs. (3.11) and (3.12) giving the desired conclusion. This we now do.

To obtain an asymptotic representation for $H_3(\eta_w)$, we expand $\exp(\sigma^3/3)$ as a power series and then make the change of variables $z = -\sigma\eta_w$. The restrictions on $\sigma$ and $\eta_w$ require $15\pi/6 < \arg(\sigma\eta_w) < 21\pi/6$, and hence our new path, $C''$, will be a ray going to infinity in the right half plane. We have upon interchanging integration and summation that

$$H_3(\eta_w) \sim \sum_{m=1}^{\infty} \frac{(-1)^m}{3^m \cdot \eta_w^{3m} \cdot m!} \int_{C''} e^{-z}\, z^{3m-1}\, dz.$$

But integration over $C''$ is equivalent to integration along the real axis from 0 to $\infty$; hence for $4\pi/3 < \arg \eta_w < 8\pi/3$

$$H_3(\eta_w) \sim \sum_{m=1}^{\infty} \frac{(-1)^m\, \Gamma(3m)}{3^m \cdot m!}\, \eta_w^{-3m}. \tag{3.9}$$

A series in ascending powers of $\eta_w$ for $H_3(\eta_w)$ may be obtained by adding and subtracting $(e^{\sigma\eta_w} - 1)e^{\sigma^3/3}/\sigma$ and $e^{\sigma}/\sigma$ to the integrand, then

$$H_3(\eta_w) = \int_{C'} \frac{e^{\sigma^3/3} - e^{\sigma}}{\sigma}\, d\sigma + \int_{C'} \frac{e^{\sigma} - e^{\sigma\eta_w}}{\sigma}\, d\sigma + \int_{C'} \frac{e^{\sigma\eta_w} - 1}{\sigma}\, e^{\sigma^3/3}\, d\sigma.$$

The first integral is a constant, and will not be evaluated. The second integral is evaluated by making the changes of variables $z = -\sigma$. The path of integration lies in the right half plane and

$$\int_{C''} \frac{e^{-z} - e^{-z\eta_w}}{z}\, dz = \int_0^{\infty} \frac{e^{-z} - e^{-z\eta_w}}{z}\, dz = \log \eta_w - \log 1.$$

Here the determination of arg (1) must be chosen appropriate to the value of $\eta_w$. To evaluate the third integral we expand $\exp(\sigma\eta_w)$ as a power series, and then interchange integration and summation which is permissible,

$$\int_{C'} \frac{e^{\sigma\eta_w} - 1}{\sigma}\, e^{\sigma^3/3}\, d\sigma = \sum_{m=1}^{\infty} \frac{\eta_w^m}{m!} \int_{C'} \sigma^{m-1} e^{\sigma^3/3}\, d\sigma$$

Making the change of variables $\sigma = -z$, and then $z = y^{1/3}$ we obtain the integral over $C'''$ of $(-1)^m 3^{1/(3m-1)} y^{1/(3m-1)} e^{-y}$. The path, $C'''$ extends to $\infty$ along a ray in the lower right half plane. Again, though, we may integrate over the real axis from 0 to $\infty$. We obtain, finally, for finite values of $\eta_w$ that

$$H_3(\eta_w) = \text{Const.} + \log \eta_w + \sum_{m=1}^{\infty} \frac{3^{\frac{1}{3}m-1}\, \Gamma(m/3)}{m!}\, (-\eta_w)^m. \tag{3.10}$$

To compare the Wasow and Tollmien solution we set $\eta_w = e^{\gamma i}\eta_T$, and determine $\gamma$ so that the sectors of validity of the asymptotic representations agree. Since $4\pi/3 < \arg \eta_w < 8\pi/3$, and $-2\pi/3 < \arg(i\eta_T) < 2\pi/3$ we obtain $\arg \gamma = \arg i + 2\pi$. As only integral powers of $\eta_w$ appear in Eqs. (3.9) and (3.10) we may set $\eta_w = i\eta_T$, also $-w_0''/w_0'$ is set equal to $+1$ for the comparison. Substituting in Eqs. (3.9) and (3.10) and comparing with Eqs. (2.11) and (2.13)

we see that $H_3(\eta_w)$ and $d\bar{\phi}_{21}/d_T - (w_0''/w_0')\log \eta_T$ are the same, since their representations for finite $\eta_T$ agree except for unevaluated constants which must be the same because of the agreement of the asymptotic representations.

Hence for finite $\eta_T$

$$\frac{dv_3}{dx} - \frac{d\phi_2^*}{dy} = \left\{C + \log \eta_T + \sum_{m=1}^{\infty} \frac{3^{m-1}\Gamma(m/3)}{m!}(-i\eta_T)^m\right\} \cdot \left\{\frac{{}_0u_3}{{}_0q_3} - B_1(y)\right\} + 0(\epsilon_T), \tag{3.11}$$

and for $|\eta_T| \to \infty$, $2\pi/3 < \arg(i\eta_T) < 2\pi/3$

$$\frac{dv_3}{dx} - \frac{d\phi_2^*}{dy} \sim \sum_{m=1}^{\infty} \frac{(-i)^m\Gamma(3m)}{3^m \cdot m!}\eta_T^{-3m}\left\{\frac{{}_0u_3}{{}_0q_3} - B_1(y)\right\} + 0(\epsilon_T). \tag{3.12}$$

It is easy to show from the power series representations of ${}_0u_3$ and ${}_0q_3$, [3], that

$$\frac{{}_0u_3}{{}_0q_3} - B_1(y) = 0(y) = 0(\epsilon\eta_T).$$

Consequently from Eqs. (3.11) and (3.12) we have the desired result that $dv_3/dx$ and $d\phi_2^*/dy$ agree to terms of $0(\epsilon)$ for finite $\eta$ and large $\eta$.

It should be mentioned that from Eq. (2.9) we have $|d\phi_2^*/dy - d\phi_2/dy| = 0(1)$ for $\eta_T$ bounded away from zero. But from the Tollmien error estimate it can be seen that $d\phi_{II}/dy$ is $0(\log \epsilon)$ and hence the percentage error in $d\phi_2/dy$ is $0(1/\log \epsilon)$. Thus the use of $\phi_2$ in computations does give the correct approximate answer.

**4. The Modified Wasow Solution.** It can be seen that the higher derivatives of the ratio ${}_0u_3/{}_0q_3$ used in defining $v_3(x, \lambda)$ have a logarithmic singularity at the origin. Precisely, $({}_0u_3/{}_0q_3)' \sim (\pi i - 2\gamma + 1) + 0(x \log x)$, and $({}_0u_3/{}_0q_3)'' \sim 0(\log x)$. This will introduce a singularity into the higher derivatives of $v_3(x, \lambda)$. It is, however, possible to construct a function $V_3(x, \lambda)$ which has the desired properties of $V_3(x, \lambda)$, but does not have this logarithmic singularity.*

If the solutions ${}_0u_3$ and ${}_0q_3$ are written as

$${}_0u_3(x) = P_1(x) + P_2(x)\log x, \qquad {}_0q_3(x) = Q_1(x) + Q_2(x)\log x,$$

the function $V_3(x, \lambda)$ is defined by

$$V_3(x, \lambda) = P_1(x) - \frac{Q_1(x)P_2(x)}{Q_2(x)} + \frac{P_2(x)}{Q_2(x)}U_3(x, \lambda). \tag{4.1}$$

Since

$$U_3(x, \lambda) = Q_1(x) + Q_2(x)\log x + \lambda^{-2/3}E(x, \lambda),$$

for all $x$, and

$$dU_3/dx = (d/dx)\{Q_1(x) + Q_2(x)\log x\} + H_3(\eta_w) + 0(\lambda^{-2/3}),$$

* This was pointed out to the author by Professor C. C. Lin.

for $\eta_w$ not zero, we obtain from Eq. (4.1) after some simple calculations that

$$V_3(x, \lambda) = {}_0u_3(x) + 0(\lambda^{-2/3}),$$

and

$$\frac{dV_3}{dx} = \frac{d_0u_3}{dx} + \frac{P_2(x)}{Q_2(x)} H_3(\eta_w) + 0(\lambda^{-2/3}).$$

It can be seen quite easily that the conclusions made concerning $v_3(x, \lambda)$ and $dv_3/dx$ in Section 3 will also hold for this new function. It seems likely that if $v_3(x, \lambda)$ is replaced by $V_3(x, \lambda)$ in [3] one could obtain rigorously the Tollmien or a better error estimate, even when $x = 0$.

## BIBLIOGRAPHY

[1] TOLLMIEN, W., "Uber die Entstehung der Turbulenz." Goettinger Nachrichten, Math-Phys. Klasse, p. 21–44 (1929).

[2] TOLLMIEN, W., "Asymptotische Integration der Stoerungsdifferentialgleichung ebener laminar Stroemungen bei hohen Reynoldschen Zahlen." Z. Angew. Math. Mech., **25**, pp. 33–50, and **27**, pp. 70–83 (1947).

[3] WASOW, W., "Asymptotic Solution of the Differential Equation of Hydrodynamics Stability in a Domain Containing a Transition Point", Ann. of Math., **58**, pp. 222–253 (1953).

[4] WASOW, W., "A study of the solutions of the differential equation $y^{(4)} + \lambda^2(xy'' + y) = 0$ for large values of $\lambda$", Ann. of Math. **52**, pp. 350–361 (1950).

[5] WATSON, G. N., *Theory of Bessel Functions*, Cambridge University Press, Second Edition (1944).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(Received February 10, 1954)

# CODING TO ACHIEVE MARKOV TYPE REDUNDANCY*

By Joseph E. Flanagan

**1. Introduction.** It is well known [6] that if one codes incoming text[1] by the device of first dividing every possible message into consecutive, nonoverlapping blocks of length $\beta$ and then applying a 0,1 code[2] to the $v^\beta$ possible constituents of a block[3], that (i) the average number of 0,1's per symbol of text is bounded from below by the entropy per symbol [6] (a constant, dependent only on the incoming text) and that (ii) one may find for any preassigned $\epsilon > 0$ a $\beta$ and a 0,1 code for the resulting block so that the average number of 0,1's per symbol of text is within $\epsilon$ of this lower bound. Roughly speaking, this latter result is achieved by taking $\beta$ large and exploiting the wealth of statistical structure attendant to such a choice of $\beta$ in determining the 0,1 sequences to be assigned to the various $\beta$ tuples (improbable $\beta$ tuples getting long codes).

Let us assume for the moment then that we have block coded the incoming text in an efficient fashion, i.e., so that the average number of 0,1's per symbol of text is near its lower bound and consequently so that $\beta$ is large. Aside from such difficulties as lag and large memory capacity that would be required under such an assumption, there arises the problem of decoding when the coded message has been subject to noise contamination. It is clear, that if for example a 1 is received as a 0, then at least a $\beta$ tuple of original text is lost ($\beta$ large)! In addition, it can be shown [6] that when incoming text is coded in this efficient fashion, then the coded text becomes close, in some sense[4], to the discrete stochastic process consisting of independent random variables each taking on the values 0,1 with probabilities $\frac{1}{2}$, $\frac{1}{2}$ respectively. Consequently, even knowing the particular 0 or 1 affected by the noise, one would be unable to make an effective guess as to the symbol in question based on the correctly received part of the coded message.

It is the purpose of this paper to study the problem of block coding incoming text in an inefficient fashion (from the noise free channel standpoint) so that the resulting 0,1 channel stochastic process is close, in some sense, to a preassigned 0,1 discrete, stationary, Markov $\alpha$ process. This could be interpreted as an attempt to combat noise in the channel by introducing redundancy in the form of Markov $\alpha$ dependency.

After describing the measure of closeness that will be used for this purpose, we shall find a necessary condition on the average length of the 0,1 code sequences used for the $\beta$ tuples, for a prescribed degree of closeness.

* This is part of work supported by the Department of the Air Force.

[1] By incoming text is meant a discrete, finite range (with $v$ elements), stationary stochastic process. In the main body of the paper we will use a more restrictive definition.

[2] That is, to every $\beta$ tuple of positive probability there will be assigned a finite sequence of 0,1's such that the sequence of 0,1's assigned to one $\beta$ tuple is not the initial segment of the sequence of 0,1's assigned to any other.

[3] Such a scheme will be referred to as $\beta$ block coding or block coding.

[4] This will be made precise in the main body of the paper.

**2. Generalization of Shannon's Entropy per Symbol.** The simplest problems of information theory are concerned with collections of messages produced from a finite set of symbols $(s_1, s_2, \cdots, s_v)$. Thus if we let the set of integers $A = (1, 2, \cdots, v)$ represent these symbols, we are led naturally to the space, $W_v$, of sequences $w$: $(\cdots, c_{-1}, c_0, c_1, \cdots)$, where $c_i \varepsilon A$ for all integers $i$.

We shall assume two probability measures $P$, $M$ defined over the smallest Borel field ($\sigma$-algebra), $F(W_v)$, of subsets of $W_v$ such that the coordinate variables $(\cdots, x_{-1}, x_0, x_1, \cdots)$ are measurable functions. Unless otherwise mentioned explicitly, we will assume the probability measures $P$, $M$ are such that the $x_n$ process is stationary with respect to both measures, and that $P$ is absolutely continuous with respect to $M$ ($P \ll M$) over any Borel field generated by a finite number of the coordinate variables.

The coordinate processes generated by the $P$ measure and the $M$ measure will be designated respectively by $(x_n, -\infty < n < \infty)_P$ and $(x_n, -\infty < n < \infty)_M$ or more compactly by $(x_n)_P$ and $(x_n)_M$. In a similar fashion expectations with respect to the $P$ measure will be written as $E_P$, etc. In the particular case when the $M$ measure defines $(x_n)_M$ to be a collection of independent random variables with $M(x_n = i) = v^{-1}$, we shall use the symbol $(x_n)_{M(v)}$ in place of $(x_n)_M$. Similarly, $(x_n)_{M(\alpha)}$ will be used in place of $(x_n)_M$ when the $M$ measure forces $(x_n)_M$ to be a Markov $\alpha$ process, that is, when

$$M(x_n = i \mid x_{n-1}, x_{n-2}, \cdots) = M(x_n = i \mid x_{n-1}, \cdots, x_{n-\alpha}) \text{ a.e.M.}$$

For any finite sample sequence $c_{t-s}, \cdots, c_t$ ($s \geqq 0$) we will have conditional probability distributions, $P(x_{t+1} = i \mid x_t = c_t, \cdots, x_{t-s} = c_{t-s})$ and $M(x_{t+1} = i \mid x_t = c_t, \cdots, x_{t-s} = c_{t-s})$ for the next symbol $x_{t+1}$. A reasonable definition of "closeness" between the processes $(x_n)_P$ and $(x_n)_{M(v)}$ would be that the two conditional probability distributions be "close" in some sense for $s$ large ($t$ fixed). If one interprets the entropy [6]

$$-\sum_{i=1}^{v} g[P(x_{t+1} = i \mid x_t = c_t, \cdots, x_{t-s} = c_{t-s})],$$

where $g(x) \equiv x \log_2 x$ and $g(0) = 0$, as a measure of "closeness" between the conditional probability distribution $P(x_{t+1} = i \mid x_t = c_t, \cdots, x_{t-s} = c_{t-s})$ and the uniform distribution $v^{-1}, \cdots, v^{-1}$ (which would be precisely the conditional probability distribution $M(x_{t+1} = i \mid x_t = c_t, \cdots, x_{t-s} = c_{t-s})$ for the case $(x_n)_{M(v)}$), we see that the number

$$\lim_{s \to \infty} E_P\{-\sum_{i=1}^{v} g[P(x_{t+1} = i \mid x_t = c_t, \cdots, x_{t-s} = c_{t-s})]\},$$

which will be independent of $t$, can be taken as a measure, in a $P$ average sense, of "closeness" between $(x_n)_P$ and $(x_n)_{M(v)}$[5]. This number, of course, has been defined by Shannon [6] as the entropy per symbol of the process $(x_n)_P$. Existence and other assorted properties of it may be found in the list of references and will be assumed to be known by the reader.

An obvious way to extend the line of reasoning mentioned above to the case

[5] This number will not be a metric.

where $(x_n)_M$ may be a Markov $\alpha$ process is to seek out a function of the probability distributions $p_1, p_2, \cdots, p_v; m_1, m_2, \cdots, m_v$, where $p_i = 0$ whenever $m_i = 0$, which will measure closeness of the two distributions in some sense, and for the case when $m_i = v^{-1}$, measure it in the same sense as described above. If we demand that this function be a continuous function and that it possess a decomposition property similar to the decomposition property valid for entropy [6, p. 19], then this function can be determined.

THEOREM 1. The only function, $f(p_1, \cdots, p_v, m_1, \cdots, m_v)$, that satisfies, for all $v$, the conditions

(i) $f$ is a continuous function

(ii) $f(p_1, \cdots, p_v, v^{-1}, \cdots, v^{-1}) = -\sum_1^v p_j \log_2 p_j - \log_2 v$

(iii) If $\sum_{s_{i-1}+1}^{s_i} p_j = q_i$, $\sum_{s_{i-1}+1}^{s_i} m_j = l_i$, $i = 1, 2, \cdots, u$, where $s_0 = 0$, $1 \leqq s_1 < s_2 < \cdots < s_u = v$,

$$\text{then } f(p_1, \cdots, p_v, m_1, \cdots, m_v) = f(q_1, \cdots, q_u, l_1, \cdots, l_u) + \sum_{j=1}^{u} q_j f\left(\frac{p_{s_{j-1}+1}}{q_j}, \cdots, \frac{p_{s_j}}{q_j}, \frac{m_{s_{j-1}+1}}{l_j}, \cdots, \frac{m_{s_j}}{l_j}\right),$$

is given by $f(p_1, \cdots, p_v, m_1, \cdots, m_v) = \sum_1^v p_j \log_2 m_j/p_j$. (We are observing the usual integration convention that $0 = 0 \log 0 = 0 \log \infty$.)

PROOF. It is easily verified that this function satisfies (i), (ii), (iii). To show the uniqueness suppose that the $m_i$, $i = 1, \cdots, v$ are rational. Then $m_i = t_i/t$, $i = 1, \cdots, v$, where $t_i, t$ are integers. Let $a_1, a_2, \cdots, a_t$ be a set of positive constants such that $\sum_1^t a_j = 1$ and $\sum_{s_{i-1}+1}^{s_i} a_j = p_i$, $i = 1, \cdots, v$, where $s_0 = 0, s_1 = t_1, s_2 = t_1 + t_2, \cdots, s_v = t$. Then by (iii) and (ii)

$$-\sum_1^t a_j \log_2 a_j - \log_2 t = f(p_1, \cdots, p_v, m_1, \cdots, m_v) + \sum_{j=1}^{v} p_j \left[-\sum_{i=1}^{t_j} g\left(\frac{a_{s_{j-1}+i}}{p_j}\right) - \log_2 t_j\right].$$

Therefore

$$f(p_1, \cdots, p_v, m_1, \cdots, m_v) = \sum_{j=1}^{v}{}' \sum_{i=1}^{t_j} a_{s_{j-1}+i} \log_2 p_j^{-1} + \sum_j p_j \log_2 \frac{t_j}{t} = \sum_1^v p_j \log_2 \frac{m_j}{p_j}.$$

In view of (i) we have proved our theorem.

In addition to items (i), (ii), (iii) of Theorem 1, it is easy to show that

(1) $\log_2 w \leqq f \leqq 0$, where $w = \min_i(m_i \mid m_i > 0)$,

(2) $f = 0$ if and only if $p_i = m_i$, $i = 1, \cdots, v$,

(3) $f = \log_2 w$ if and only if there exists and $i$, say $i_0$, such that $p_{i_0} = 1$ and $m_{i_0} = w$.

The proofs for statements (1), (2), and (3) are left to the reader.

In all this work we shall interpret the $M$ distribution as the fixed standard of comparison in the same sense that the uniform distribution serves as the standard of comparison in information theory. It should be pointed out that

$f$ cannot be interpreted as a metric since for example the symmetry property is lacking.

THEOREM 2. If the process $(x_n)_M$ is Markov $\alpha$, then the random variables

$$\sum_{j=1}^{v} P(x_{t+1} = j \mid x_t, x_{t-1}, \cdots, x_{t-s}) \log_2 \frac{M(x_{t+1} = j \mid x_t, x_{t-1}, \cdots, x_{t-s})}{P(x_{t+1} = j \mid x_t, x_{t-1}, \cdots, x_{t-s})},$$

$s = \alpha - 1, \alpha, \alpha + 1, \cdots$, form a lower semi-martingale with respect to the Borel fields generated by $x_t, \cdots, x_{t-s}$, $s = \alpha - 1, \alpha, \alpha + 1, \cdots$, and these random variables converge mean $P$ and a.e.P to the limiting function

$$\sum_{j=1}^{v} P(x_{t+1} = j \mid x_t, x_{t-1}, \cdots) \log_2 \frac{M(x_{t+1} = j \mid x_t, x_{t-1}, \cdots)}{P(x_{t+1} = j \mid x_t, x_{t-1}, \cdots)}.$$

PROOF. Since $P(x_{t+1} = j \mid x_t, \cdots, x_{t-s})$, $s = 0, 1, \cdots$, is a martingale with respect to the Borel fields generated by $x_t, \cdots, x_{t-s}$, $s = 0, 1, \cdots$ and since $x \log_2 x$ is continuous and convex in the unit interval we have that $-g[P(x_{t+1} = j \mid x_t, \cdots, x_{t-s})]$, $s = 0, 1, \cdots$, forms a lower semi-martingale with respect to the Borel fields generated by $x_t, \cdots, x_{t-s}$, $s = 0, 1, \cdots$. (The functions $-g[P(x_{t+1} = j \mid x_t, \cdots, x_{t-s})]$, $s = 0, 1, \cdots$, are obviously $P$ integrable, in fact by Jensen's inequality they are uniformly bounded from above by $\log_2 v$.) Now $P(x_{t+1} = j \mid x_t, \cdots, x_{t-s}) \log_2 M(x_{t+1} = j \mid x_t, \cdots, x_{t-\alpha+1})$, $s = \alpha - 1$, $\alpha, \alpha + 1, \cdots$, forms a martingale with respect to the Borel fields generated by $x_t, \cdots, x_{t-s}$, $s = \alpha - 1, \alpha, \alpha + 1, \cdots$, since for $s \geqq \alpha - 1$

$$\int_{x_t = c_t, \cdots, x_{t-s} = c_{t-s}} P(x_{t+1} = j \mid x_t, \cdots, x_{t-s-1}) \log_2 M(x_{t+1} = j \mid x_t, \cdots, x_{t-\alpha+1})\, dP =$$

$$\sum_{c_{t-s-1}} P(x_{t+1} = j, x_t = c_t, \cdots, x_{t-s-1} = c_{t-s-1}) \log_2 M(x_{t+1} = j \mid x_t = c_t, \cdots, x_{t-\alpha+1} = c_{t-\alpha+1})$$

$$= \int_{x_t = c_t, \cdots, x_{t-s} = c_{t-s}} P(x_{t+1} = j \mid x_t, \cdots, x_{t-s}) \log_2 M(x_{t+1} = j \mid x_t, \cdots, x_{t-\alpha+1})\, dP.$$

Using the fact that a finite sum of lower semi-martingales is a lower semi-martingale (assuming the Borel fields concerned are compatible, which is the case here) we have proved the first part of Theorem 2.

To prove the second part of the theorem we note that convergence a.e.P of the indicated functions is obvious since

$$\lim_{s \to \infty} P(x_{t+1} = j \mid x_t, \cdots, x_{t-s}) = P(x_{t+1} = j \mid x_t, x_{t-1}, \cdots) \text{ a.e.P.}$$

Since by (1) we have that the indicated functions ($s \geqq \alpha - 1$) are bounded from below by

$$\log_2 \{\min_{j, c_t, \cdots, c_{t-\alpha+1}} [M(x_{t+1} = j \mid x_t = c_t, \cdots, x_{t-\alpha+1} = c_{t-\alpha+1}) > 0]\},$$

we have convergence mean $P$ by Lebesgue's theorem.

We are now in a position to define a measure of closeness, $H[(x_n)_P, (x_n)_{M(\alpha)}]$, between the processes $(x_n)_P$ and $(x_n)_{M(\alpha)}$, i.e., let

$$H[(x_n)_P, (x_n)_{M(\alpha)}] = \lim_{s\to\infty} E_P\left[\sum_{j=1}^{r} P(x_{t+1} = j \mid x_t, \cdots, x_{t-s}) \log_2 \frac{M(x_{t+1} = j \mid x_t, \cdots, x_{t-s})}{P(x_{t+1} = j \mid x_t, \cdots, x_{t-s})}\right]. \tag{4}$$

The limit $H[(x_n)_P, (x_n)_{M(\alpha)}]$ exists by Theorem 2 and is independent of $t$ by the stationarity assumption on $(x_n)_P$ and $(x_n)_{M(\alpha)}$. If one designates by $H[(x_n)_P]$ the entropy per symbol of the process $(x_n)_P$, then we have from (4), in view of the Markov property,

$$H[(x_n)_P, (x_n)_{M(\alpha)}] = H[(x_n)_P] + E_P[\textstyle\sum_{j=1}^{r} P(x_{t+1} = j \mid x_t, \cdots, x_{t-\alpha+1}) \cdot \log_2 M(x_{t+1} = j \mid x_t, \cdots, x_{t-\alpha+1})]. \tag{5}$$

Many of the important theorems of information theory, as for example the limit theorem [5, p. 212] that states that $(-\log p(x_1, \cdots, x_n))/n$ converges mean $P$ as $n$ approaches infinity and to the limit $H$ if the $(x_n)_P$ process is ergodic, have obvious analogues [3], with corresponding proofs, for this slightly more general definition of entropy per symbol.

**3. Block Coding.** As mentioned earlier in this paper we desire to code (0,1 code) incoming text[6] (discrete, stationary, ergodic, Markov $\beta$ chain) in such a fashion that the resulting coded process (0,1 stationary, ergodic, process) will be close, that is, $H[(x_n)_P, (x_n)_{M(\alpha)}]$ is near 0, to some preassigned stationary 0,1 ergodic Markov $\alpha$ chain.

If the coded process is near that of the preassigned Markov $\alpha$ chain, in this $H[(x_n)_P, (x_n)_{M(\alpha)}]$ sense, then on the average we will be able to predict a missing channel symbol given the preceding channel sequence on the basis of the transition probabilities of the preassigned Markov $\alpha$ chain.

The mathematical model that will be used for this purpose is given as follows. We suppose the incoming text is representable as a discrete, stationary, ergodic, Markov $\beta$ chain, $(y_n, -\infty' < n < \infty)$. The assumption of ergodicity is not too unreasonable in view of the fact that a stationary, Markov $\beta$ chain, $(y_n)$, is ergodic if and only if the associated stationary, Markov vector chain, $(\bar{y}_n)$,

$$\bar{y}_n = (y_n, \cdots, y_{n+\beta-1})$$

has one ergodic class [1, p. 460], that is, every $\beta$ tuple $(y_n, \cdots, y_{n+\beta-1})$ can be reached from any other $\beta$ tuple in a finite number of steps with positive probability. We shall use a block coding device, that is, we shall code successive non-overlapping blocks of $\beta$ tuples of the original process $(y_n)$. The $\beta$ tuple process, $(z_n, -\infty < n < \infty)$, where $z_n = \bar{y}_{n\beta}$, will clearly be a stationary Markov chain

[6] Note that we have made the definition of incoming text more restrictive than the definition given earlier.

since $(\bar{y}_n)$ is a stationary Markov chain. We will further assume that it is ergodic (this would be assured if we assumed that the Markov chain $(\bar{y}_n)$ is aperiodic).

We shall now code each $\beta$ tuple of positive probability into a finite sequence of 0,1's such that the 0,1 sequence for one tuple is not the initial segment of the 0,1 sequence of any other. For any such assignment of 0,1 sequences we will say that we have $\beta$ block coded the incoming text. Clearly the resulting 0,1 coded process need not be stationary or Markov. What we desire is to block code the incoming text so that the resulting coded process is stationary and near, in the $H[(x_n)_P, (x_n)_{M(\alpha)}]$ sense, to the preassigned 0,1 stationary, Markov $\alpha$ chain and such that the average length of the 0,1 code sequences is kept within reason. If we designate by $H[(y_n)]$ the entropy per symbol of the $(y_n)$ process, then clearly $H[(z_n)]$, the entropy per symbol of the $(z_n)$ process, is

$$H[(z_n)] = \beta H[(y_n)] \qquad (\beta > 0). \tag{6}$$

Let us put the $\beta$ tuples of positive probability in 1-1 correspondence with the integers $1, 2, \cdots, k$, with $\bar{p}_1, \bar{p}_2, \cdots, \bar{p}_k$ as the corresponding probability distribution. Let $l_i$ represent the length (number of 0,1's) of the 0,1 sequence representing the $i^{\text{th}}$ $\beta$ tuple. The average length of the code per $\beta$ tuple is then $\sum_1^k \bar{p}_i l_i$. We shall now concern ourselves with the structure of the coded 0,1 process for various choices of codes. To do this let us suppose that the various 0's can be distinguished from one another and similarly that the various 1's can be distinguished from one another.[7] That is, the $k$ sequences of 0,1's will be interpreted as $k$ sequences of distinct elements, $a_{ij}$'s, as follows:

$$\begin{aligned} &\bar{p}_1 : a_{11}a_{12} \cdots a_{1l_1} \\ &\bar{p}_2 : a_{21}a_{22} \cdots a_{2l_2} \\ &\cdots\cdots\cdots\cdots \\ &\bar{p}_k : a_{k1}a_{k2} \cdots a_{kl_k}, \end{aligned} \tag{7}$$

where the $\bar{p}_1, \cdots, \bar{p}_k$ on the left represent the probabilities of the corresponding $\beta$ tuples. We have then a natural matrix of transition probabilities between the various $a_{ij}$'s. That is,

$$\begin{aligned} P(a_{ij} \mid a_{kl}) &= 1, && \text{if } i = k, j = l + 1 \\ &= p_k(i), && \text{if } l = l_k, j = 1 \\ &= 0 && \text{otherwise,} \end{aligned} \tag{8}$$

where $p_k(i)$ is the probability of the $i^{\text{th}}$ $\beta$ tuple given the $k^{\text{th}}$ $\beta$ tuple, that is, the transition probabilities for the $(z_n)$ Markov chain. Consequently any resulting 0,1 coded process can be represented as a function of a Markov chain with $l_1 + l_2 + \cdots + l_k$ states.

Since the stationary Markov chain $(z_n)$ is assumed ergodic (and consequently

[7] By interpreting the various $a_{ij}$'s as states of a Markov chain, this device will give us an underlying associated Markov process.

has only one ergodic class) we see that the Markov chain with (8) as its matrix of transition probability will have only one ergodic class. We shall now determine the stationary absolute probabilities, $q_{ij}$, associated with the matrix of transition probabilities, (8). The condition to be satisfied is as follows:

$$q_{i1} = q_{1l_1} p_1(i) + q_{2l_2} p_2(i) + \cdots + q_{kl_k} p_k(i), \qquad i = 1, 2, \cdots, k$$

(9)

$$q_{i1} = q_{i2} = \cdots = q_{il_i}, \qquad i = 1, 2, \cdots, k.$$

That is,

$$q_{il_i} = q_{1l_1} p_1(i) + q_{2l_2} p_2(i) + \cdots + q_{kl_k} p_k(i), \quad i = 1, 2, \cdots, k$$

(10)

$$q_{i1} = q_{i2} = \cdots = q_{il_i}, \qquad i = 1, 2, \cdots, k.$$

In view of (8) and the fact that $p_i(j)$ represents the stationary transition probabilities of the stionary Markov chain $(z_n)$, we see that

$$q_{il_i} = c\bar{p}_i, \qquad \sum_1^k l_i c\bar{p}_i = 1.$$

This gives us that

$$q_{i1} = q_{i2} = \cdots = q_{il_i} = \frac{\bar{p}_i}{\sum_i l_i \bar{p}_i}, \qquad i = 1, 2, \cdots, k. \tag{11}$$

The stationary, ergodic, Markov chain with states $a_{ij}$, transition matrix (8) and stationary absolute probabilities (11) will be referred to as the associated Markov chain for the code (7), and will be designated by $(a_n, -\infty < n < \infty)$ or more briefly $(a_n)$.

Since the associated Markov chain $(a_n)$ is stationary, the process $[f(a_n), -\infty < n < \infty]$ is stationary, where $f$ is arbitrary. Henceforth we shall assume that $f$ has range 0,1 and that it maps the $k$ sequences of $a_{ij}$'s given by (7) into $k$ sequences of 0,1's that satisfy the initial segment condition[2]. Note that we must start the 0,1 coded process, by this device, in accordance with (11) to obtain stationarity.

**4. Main Theorem.** We shall now show that the $\beta$ block coding scheme (if any) of smallest average length such that $H[f(a_n), (x_n)_{M(\alpha)}]$ is within $+d^2$ of 0, where $+d^2$ is sufficiently small, must be one of the finite set of $\beta$ block codes of average length $\leqq (\beta H[(y_n)]/-d^2 - \log_2 \bar{m})$.

THEOREM 3. Let $(x_n, -\infty < n < \infty)_{M(\alpha)}$ be a 0,1 stationary Markov $\alpha$ chain and let $M$ represent the probability measure associated with it. Then

(i) $H[f(a_n), (x_n)_{M(\alpha)}] = (\beta H[(y_n)]/\sum_i l_i \bar{p}_i) + \sum_{i_1, \cdots, i_{\alpha+1}} P[f(a_1) = i_1, \cdots, f(a_{\alpha+1}) = i_{\alpha+1}] \log_2 M(x_{\alpha+1} = i_{\alpha+1} | x_\alpha = i_\alpha, \cdots, x_1 = i_1)$,

(ii) A necessary condition for a $\beta$ block coding scheme to exist and be such that $H[f(a_n), (x_n)_{M(\alpha)}] = -d^2$, where $-d^2 > \log_2 \bar{m}$ and

$$\bar{m} \equiv \max_{i_{\alpha+1}, \cdots, i_1} [M(x_{\alpha+1} = i_{\alpha+1} \mid x_\alpha = i_\alpha, \cdots, x_1 = i_1)],$$

is that

$$\sum_1^k l_i \bar{p}_i \leqq \frac{\beta H[(y_n)]}{-d^2 - \log_2 \bar{m}} = \frac{H[(z_n)]}{-d^2 - \log_2 \bar{m}}.$$

PROOF. To prove (i) we need the following lemma.

LEMMA 1. The entropy per symbol, $H[f(a_n)]$, of the stationary 0,1 process $[f(a_n), -\infty < n < \infty]$ is precisely the same as the entropy per symbol, $H[(a_n)]$, of the stationary Markov chain $(a_n, -\infty < n < \infty)$, that is,

$$H[f(a_n)] = H[(a_n)] = \frac{\beta H[(y_n)]}{\sum_i l_i \bar{p}_i}.$$

PROOF OF LEMMA. Let $i_1, \cdots, i_n$ be a sequence of 0,1's drawn from the process $[f(a_n), -\infty < n < \infty]$ and let $N(i_1, \cdots, i_n)$ be the number of corresponding sequences of $a_{ij}$'s, of length $n$, that could conceivably map into $i_1, \cdots, i_n$ under $f$. We shall now show that $N(i_1, \cdots, i_n)$ is bounded from above by a constant independent of $n$.

Given a sequence, $i_1, \cdots, i_n$, of 0,1's drawn from $[f(a_n)]$, let us choose a terminal segment of an $l_i$ tuple of the $a_{ij}$'s (7) that maps into an initial segment of $i_1, \cdots, i_n$, say into the first $k$ of them, $i_1, \cdots, i_k$. Then there can exist at most one other $l_i$ tuple of the $a_{ij}$'s that maps under $f$ into an initial segment of $i_{k+1}, \cdots, i_n$. For if there exists two or more such $l_i$ tuples of $a_{ij}$'s, then the corresponding 0,1 sequences under $f$ of these $a_{ij}$ sequences would violate the initial segment condition[2]. In a similar fashion, there can exist at most one $l_i$ tuple of $a_{ij}$'s that will map into an initial segment of the remaining 0,1's. If one continues this process until there remains only an end piece of the $i_1, \cdots, i_n$ sequence smaller than the largest of the $l_i$, $i = 1, 2, \cdots, k$, it is obvious that there can be at most $k$ possible ways of finding corresponding $a_{ij}$ tuples. Consequently, we have shown that for the particular terminal segment that we began this whole affair with, there are at most $k$ possible sequences, of length $n$, of $a_{ij}$'s beginning with this terminal segment that map into the given sequence $i_1, \cdots, i_n$. Since there are at most $l_1 + l_2 + \cdots + l_k$ possible terminal segments that can map into an initial segment of the given sequence, $i_1, \cdots, i_n$, we have $(l_1 + l_2 + \cdots + l_k)k$ as an upper bound on $N(i_1, \cdots, i_n)$, that is,

$$N(i_1, \cdots, i_n) \leqq (l_1 + l_2 + \cdots + l_k)k. \tag{12}$$

Using for the moment the notation of [6] we have[3]

$$H(\xi, \eta) = H(\xi) + H_\xi(\eta),$$

where $\xi$ denotes the (universe of) sequences $f(a_i)$ of length $n$, and $\eta$ denotes the (universe of) sequences $(a_i)$ of length $n$. Because of the functional relation $a_i \to f(a_i)$ we have also $H(\xi, \eta) = H(\eta)$. Because of (12) we have

$$H_\xi(\eta) \leqq \log_2(l_1 + l_2 + \cdots + l_k)k.$$

[3] The following paragraph was pointed out by the referee to be a simplification of the author's original proof.

Hence

$$\lim_{n\to\infty} \frac{1}{n} H(\eta) = \lim_{n\to\infty} \frac{1}{n} H(\xi),$$

which proves the lemma.

The entropy per symbol, $H[(a_n)]$, of the Markov chain $(a_n)$ is easily calculated from (8) and (11) to be

$$H[(a_n)] = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\bar{p}_i}{\sum_i l_i \bar{p}_i} p_i(j) \log_2 p_i(j) = \frac{H[(z_n)]}{\sum_i l_i \bar{p}_i} = \frac{\beta H[(y_n)]}{\sum_i l_i \bar{p}_i}. \tag{13}$$

This in conjunction with (5) and Lemma 1 gives us item (i) of Theorem 3. Item (ii) of Theorem 3 follows immediately from (i)[9].

**5. General Remarks.** It is usually natural to suppose that the comparison process $(x_n)_M$, in addition to the above mentioned properties, is ergodic (metrically transitive). This would always be the case, for example, when the process $(x_n)_M$ consists of independent random variables. We shall now show that this forces $M$ to be absolutely continuous with respect to $P$, or $M$ to be singular with respect to $P$ ($M \ll P$ or $M \perp P$). If in addition, we suppose that the $(x_n)_P$ process is ergodic, we shall have $P = M$ or $P \perp M$. Thus singular probability measures play a key role in the typical situations met in this field.

THEOREM 4. Let $P$ and $M$ be totally finite measures defined over a Borel field, $F(W)$, of subsets of an abstract space $W$. Let $T$ be a 1–1 point transformation defined over $W$, such that $T$ is measure preserving, $P$, $M$, for all elements of $F(W)$. (Assume also that $P(W) > 0$, $M(W) > 0$.)

(i) If $P \ll M$ and if $T$ is ergodic with respect to the $M$ measure, then $P = (P(W)/M(W))M$.

(ii) If $T$ is ergodic with respect to say the $M$ measure, then either $M \ll P$ or $P \perp M$.

PROOF. The proof of (i) is trivial since if $Q \varepsilon F(W)$, then

$$P(Q) = \int_Q f\, dM = \int_{TQ} f\, dM = \int_{TQ} Tf\, dM,$$

where $f$ is the associated density function. Therefore $f = Tf$ a.e.M. Ergodicity on $M$ now forces $f$ to be a constant a.e.M. This proves (i).

To prove (ii) we use the Lebesgue decomposition theorem to represent $P$ as $P = P_c + P_s$, where $P_c$ and $P_s$ are totally finite measures over $F(W)$ such that $P_c \ll M$ and $P_s \perp M$. We shall first show that the transformation $T$ is measure preserving with respect to both $P_c$ and $P_s$. To do this we use the fact that $P \ll P + M$. Consequently there exists a measurable function $f$ such that for every $Q \varepsilon F(W)$ we have

$$P(Q) = \int_Q f\, d(P + M).$$

[9] See [2], [4] for further discussions of codes.

Recalling that for every $Q \varepsilon F(W)$, $P_s(Q) = P(Q \cap A)$ and $P_c(Q) = P(Q \cap B)$, where

$$B = (w \mid 0 \leqq f(w) < 1), \qquad A = (w \mid f(w) = 1), \tag{14}$$

we see that one need only show that $A$ and $B$ are invariant sets under $T$ (or they differ from their transforms by a set of $P$ measure 0). But this is trivial, since $T$ being measure preserving with respect to the $P$ and $M$ measures, gives us that

$$P(Q) = \int_Q f\, d(P + M) = \int_{TQ} f\, d(P + M) = \int_{TQ} Tf\, d(P + M).$$

Therefore $f = Tf$ a.e.$P + M$. In particular then, $f = Tf$ a.e.P. This gives us, in view of (14), the desired result.

Now since $M$ is ergodic and $P_c \ll M$, we see by (i) and the above result, that $P_c = (P_c(W)/M(W))M$ ($P_c(W)$ may possibly be 0). Therefore

$$P = \frac{P_c(W)}{M(W)}\, M + P_s\,. \tag{15}$$

We see immediately from (15) that if $P_c(W) > 0$, then $M \ll P$. If $P_c(W) = 0$, then $P = P_s \perp M$. This proves Theorem 4.

Another remark is in order here concerning the starting probabilities (11). It is trivial to construct examples of block codes such that the resulting coded process $[f(a_n)]$ is stationary and yet the starting probabilities for the Markov chain $(a_n)$ are different than (11). In fact, in practice it would be desirable to start out the process $(a_n)$ with one of the symbols $a_{11}, \cdots, a_{k1}$. The following theorem shows that no matter how one starts out the Markov chain $(a_n)$ to get stationarity for the process $[f(a_n)]$, the resulting stationary process $[f(a_n)]$ will have the same statistical structure.

THEOREM 5. Let $[f(a_n), -\infty < n < \infty]_q$ be a stationary 0,1 coded process resulting from some particular choice of starting probabilities $q_{ij}$. If

$$[f(a_n), -\infty < n < \infty]$$

is the stationary process resulting from the starting probabilities

$$q_{a1} = q_{a2} = \cdots = q_{ai} = \frac{\bar{p}_i}{\sum_i l_i \bar{p}_i}, \qquad i = 1, 2, \cdots, k,$$

then

$$P[f(a_n) = i_n, \cdots, f(a_{n+t}) = i_{n+t}] = P_q[f(a_n) = i_n, \cdots, f(a_{n+t}) = i_{n+t}]$$

for all $n, t$, where $P_q$, $P$ represent the probability measures identified with $[f(a_n), -\infty < n < \infty]_q$, $[f(a_n), -\infty < n < \infty]$ respectively.

PROOF. Let $(a_n, -\infty < n < \infty)_q$ represent the $a_{ij}$ state Markov chain defined by the starting probabilities (11). It will have only one ergodic class since $(z_n)$ has only one ergodic class. Consequently $P_q(a_n = a_{ij})$ converges $(C, 1)$, as

$n \to \infty$, to a limit independent of the starting probabilities. Therefore

$$P_q[f(a_n) = i_1]$$

converges $(C, 1)$, as $n \to \infty$, to a limit independent of the starting probabilities $q_{ij}$. But $[f(a_n), -\infty < n < \infty]_q$ is assumed to be stationary, consequently $P_q[f(a_n) = i_1]$ is independent of $n$ for any choice of $q_{ij}$. In view of the last two sentences we must have that $P_q[f(a_n) = i_1]$ is a constant and independent of $q_{ij}$ and $n$. In particular then,

$$P_q[f(a_n) = i_1] = P[f(a_n) = i_1].$$

Using precisely the same reasoning as mentioned above one can show that in general $P_q[f(a_n) = i_n, \cdots, f(a_{n+t}) = i_{n+t}]$ is independent of $q_{ij}$. As seen by the proof, this theorem is applicable to any two stationary processes

$$[f(a_n), -\infty < n < \infty]_q$$

and $[f(a_n), -\infty < n < \infty]_{\bar{q}}$ where the Markov chains $(a_n, -\infty < n < \infty)_q$ and $(a_n, -\infty < n < \infty)_{\bar{q}}$ are ergodic and $q$, $\bar{q}$ represent different choices of starting probabilities.

## REFERENCES

[1] Doob, J. L., *Stochastic Processes*, Wiley, (1953), pp. 459–460.

[2] Fano, R. M., *Technical Report No. 65*, The Research Laboratory of Electronics, M.I.T. March 17, 1949.

[3] Flanagan, J. E., *Topics in Information Theory*, doctoral thesis, University of Illinois, 1953.

[4] Huffman, D. A., *A Method for the Construction of Minimum Redundancy Codes*, Proceedings of the I.R.E., **40**, no. 9, 1952.

[5] McMillan, Brockway, *The Basic Theorems of Information Theory*, Annals of Mathematical Statistics, Vol. 24, pp. 196–219.

[6] Shannon, C., and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.

Carnegie Institute of Technology

(Received January 7, 1954)

# AN APPLICATION OF SOMMERFELD'S COMPLEX ORDER WAVE FUNCTIONS TO ANTENNA THEORY*

By Charles H. Papas

**Introduction.** Suppose we desire to construct a solution to the wave equation

$$(\nabla^2 + k^2)u = 0 \tag{1}$$

where $u$ must satisfy the Sommerfeld radiation condition and must assume prescribed values on the surface of a sphere of radius $a$. Using spherical coordinates $r$, $\theta$, $\phi$ with origin at center of sphere, assuming that $u$ is independent of the azimuthal angle $\phi$, and requiring that on the surface of sphere

$$\alpha u + \beta \partial u/\partial r = f(\theta) \tag{2}$$

where $\alpha$ and $\beta$ are constants, we consider a solution to (1) of the form

$$u(r, \theta) = \sum_{n=0}^{\infty} A_n P_n(\cos\theta) h_n^{(1)}(kr) \tag{3}$$

where $P_n(\cos\theta)$ is the Legendre *polynomial* of *integral order* $n$ and $h_n^{(1)}(kr)$ is the spherical Hankel function of the same order. The $A_n$'s are constants which have to be determined by requiring that (3) satisfy (2) when $r = a$. As it stands, (3) satisfies the radiation condition,

$$\lim_{r\to\infty} r(\partial u/\partial r - iku) = 0 \tag{4}$$

because as $r \to \infty$ the spherical Hankel functions have the behaviour of an outwardly propagating wave,.

$$h_n^{(1)}(kr) \sim (-i)^{n+1} e^{ikr}/kr \tag{5}$$

To determine the unknown constants, the $A_n$'s, we substitute (3) into (2) and use the *orthogonality* property of the Legendre polynomials,

$$\int_0^{\pi} P_n(\cos\theta)\, P_m(\cos\theta) \sin\theta\, d\theta = \delta_{nm}\,[2/(2m+1)] \tag{6}$$

where $\delta_{nm} = 0, 1$ when $n \neq m$, $n = m$ respectively. Thus

$$A_m = \frac{2m+1}{2} \frac{\displaystyle\int_0^{\pi} f(\theta) P_m(\cos\theta) \sin\theta\, d\theta}{[\alpha h_m^{(1)}(ka) + \beta(\partial/\partial r) h_m^{(1)}(ka)]} \tag{7}$$

Substituting (7) into (3) we have the desired solution. It is clear that the crux of this problem is the orthogonality of the Legendre polynomials. If this orthogonality did not exist, it would be impossible to determine the $A_n$'s. However, for certain problems it would be much more convenient and "natural" if the radial functions, instead of the angular functions, were orthogonal. This is true, for example, if the wave function has to satisfy on a conical surface bound-

* This work was supported by the Office of Naval Research.

ary conditions which depend on $r$. In the boss antenna problem the wave function must satisfy just such boundary conditions.

In this note we first develop some properties of the complex order wave functions.[1] We find that $u$ can be represented in the form of an infinite series, each term of which is the product of an angular function and a radial function. The angular and radial functions prove to be the Legendre functions and the spherical Hankel functions of *complex order* respectively. And now it is the spherical Hankel functions of complex order that are orthogonal! Then we apply these wave functions to the calculation of the input impedance of a boss antenna.

**Complex Order Wave Functions.** Let us consider in spherical coordinates $r, \theta, \phi$ a $TM$ field, $\mathbf{E} = \mathbf{r}E_r + \boldsymbol{\theta}E_\theta$, $\mathbf{H} = \boldsymbol{\phi}H_\phi$ where $\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi}$, are the unit vectors and $E_r$, $E_\theta$, $H_\phi$ are the non-vanishing components of the electromagnetic field. For steady-state, i.e., time dependence $\exp(-i\omega t)$, the complex vectors $\mathbf{E}$ and $\mathbf{H}$ satisfy the Maxwell equations,

$$\nabla \times \mathbf{E} = i\omega\mu\mathbf{H} \tag{8}$$

$$\nabla \times \mathbf{H} = -i\omega\mathbf{E} \tag{9}$$

from which it follows that

$$\nabla \times \nabla \times \mathbf{H} = k^2\mathbf{H} \tag{10}$$

where $k^2 = \omega^2\epsilon\mu$ and $\epsilon$, $\mu$ denote the dielectric constant and permeability of free space. If we assume that $\mathbf{E}$ and $\mathbf{H}$ are independent of the azimuthal angle $\phi$, then in spherical coordinates (10) takes the form

$$\frac{\partial^2}{\partial r^2}(rH_\phi) + \frac{1}{r}\frac{\partial}{\partial\theta}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(H_\phi \sin\theta)\right] + k^2 rH_\phi = 0 \tag{11}$$

If we set

$$H_\phi = i\omega\epsilon\partial u/\partial\theta \tag{12}$$

then (11) can be written as

$$i\omega\epsilon(\partial/\partial\theta)[(\nabla^2 + k^2)u(r, \theta)] = 0 \tag{13}$$

We choose $u(r, \theta)$ such that

$$(\nabla^2 + k^2)u(r, \theta) = 0 \tag{14}$$

for all permissible values of $r$ and $\theta$. Then according to (9) and (12) the electromagnetic field is given by the following expressions:

$$E_r = -\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right) \tag{15}$$

$$E_\theta = \frac{1}{r}\frac{\partial}{\partial\theta}\frac{\partial}{\partial r}(ru) \tag{16}$$

$$H_\phi = i\omega\epsilon\frac{\partial}{\partial\theta}u \tag{17}$$

Now we examine the solution of the wave equation (14) that satisfies the radiation condition (4) and the boundary condition,

$$\partial(ru)/\partial r = 0 \qquad \text{when} \qquad r = a \tag{18}$$

In view of (16) it is seen that the condition (18) is equivalent to requiring that $E_\theta$ vanish over surface of sphere of radius $a$.

Writing $u(r, \theta)$ explicitly in separated form, $u = R(r)P(\cos\theta)$, (14) yields two ordinary differential equations

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP}{d\theta}\right) + CP = 0 \tag{19}$$

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + k^2r^2 = C \tag{20}$$

where $C$ is the separation constant. If we choose $C = \nu(\nu + 1)$ where $\nu$ is unrestricted (a complex number), the solution of (19) is the *Legendre function* $P_\nu(\cos\theta)$. This function is identical to the hypergeometric function, i.e.,

$$P_\nu(\cos\theta) = F(-\nu, \nu + 1, 1, \tfrac{1}{2} - \tfrac{1}{2}\cos\theta) \tag{21}$$

and is finite for all $\theta$ except $\theta = \pi$. The corresponding solution of (20) is the spherical Hankel function of the first kind $h_\nu^{(1)}(kr)$ and it is related to the cylindrical Hankel function of the first kind $H_{\nu+\frac{1}{2}}^{(1)}(kr)$, according to the relation,

$$h_\nu^{(1)}(kr) = (\pi/2kr)^{\frac{1}{2}}H_{\nu+\frac{1}{2}}^{(1)}(kr) \tag{22}$$

Hence, we can write the solution of (14) as

$$u(r, \theta) = \sum A_\nu h_\nu^{(1)}(kr)P_\nu(\cos\theta) \tag{23}$$

where the summation is over all values of $\nu$ determined by the boundary condition (18). Substituting (23) into (18) we obtain

$$\sum A_\nu(\partial/\partial r)(ah_\nu^{(1)}(ka)P_\nu(\cos\theta) = 0 \tag{24}$$

To satisfy (24) the $\nu$'s are chosen so that

$$[(\partial/\partial r)(rh_\nu^{(1)}(kr))]_{r=a} = 0 \tag{25}$$

We denote these $\nu$'s as $\nu_1$, $\nu_2$, $\nu_3$, etc.

The radial functions $h_\nu^{(1)}(x)$ are orthogonal over the range $x = ka$ to $x = \infty$. To show this we recall from (20) that for any $\nu$, say, $\nu_n$

$$x(d^2/dx^2)(xh_{\nu_n}^{(1)}(x)) + (x^2 - \nu_n(\nu_n + 1))h_{\nu_n}^{(1)}(x) = 0 \tag{26}$$

and for any other value of $\nu$, say, $\nu_m$

$$x(d^2/dx^2)(xh_{\nu_m}^{(1)}(x)) + (x^2 - \nu_m(\nu_m + 1))h_{\nu_m}^{(1)}(x) = 0 \tag{27}$$

Multiplying (26) by $h_{\nu_m}^{(1)}$ and (27) by $h_{\nu_n}^{(1)}$ and then integrating the difference from $x = ka$ to $x = \infty$, we obtain

$$[\nu_m(\nu_m + 1) - \nu_n(\nu_n + 1)] \int_{ka}^{\infty} h_{\nu_n}^{(1)}(x) h_{\nu_m}^{(1)}(x)\, dx$$

$$= \int_{ka}^{\infty} \left[ x h_{\nu_n}^{(1)}(x) \frac{d^2}{dx^2} (x h_{\nu_m}^{(1)}(x)) - x h_{\nu_m}^{(1)}(x) \frac{d^2}{dx^2} (x h_{\nu_n}^{(1)}(x)) \right] dx \tag{28}$$

$$= x h_{\nu_n}^{(1)}(x) \frac{d}{dx} (x h_{\nu_m}^{(1)}(x) - x h_{\nu_m}^{(1)}(x) \frac{d}{dx} (x h_{\nu_n}^{(1)}(x)) \Big]_{ka}^{\infty}$$

where the last equality results from an integration by parts. The integrated term disappears at $x = ka$ and $x = \infty$ by virtue of (25) and the asymptotic behaviour of $h_\nu^{(1)}(x)$. Thus

$$\int_{ka}^{\infty} h_{\nu_n}^{(1)}(x) h_{\nu_m}^{(1)}(x)\, dx = \delta_{nm} N_{\nu_n}(ka) \tag{29}$$

$N_{\nu_n}(ka)$ is a normalization factor which can be obtained from (28) by an application of de l'Hospital's rule for the limit $\nu_n \to \nu_m$, and, of course, $\delta_{nm}$ is the well-known Kronecker delta.

Hence, the solution of (14) which satisfies the boundary condition (18) and the Sommerfeld radiation condition is given by a sum over all $\nu$'s:

$$u(r, \theta) = \sum A_\nu h_\nu^{(1)}(kr) P_\nu(\cos\theta) \tag{30}$$

where the $A_\nu$'s are as yet undetermined constants. Once we are given $u(r, \theta_0)$ on any conical surface $\theta = \theta_0$, the $A_\nu$'s are determinable by virtue of the orthogonality condition (29). Contrasting (3) and (30) we see that (3) is appropriate for boundary conditions on a sphere, whereas (30) is appropriate for cones.

Substituting (30) into (15), (16), and (17) we find the electromagnetic field in terms of the complex order wave functions, $h_\nu^{(1)}(kr) P_\nu(\cos\theta)$,

$$H_\phi(r, \theta) = i\omega\epsilon \sum A_\nu h_\nu^{(1)}(kr)(\partial/\partial\theta) P_\nu(\cos\theta), \tag{31}$$

$$E_\theta(r, \theta) = \frac{1}{r} \sum A_\nu (\partial/\partial r)(r h_\nu^{(1)}(kr))(\partial/\partial\theta) P_\nu(\cos\theta), \tag{32}$$

and since $P_\nu(\cos\theta)$ satisfies (19) when $C = \nu(\nu + 1)$,

$$E_r(r, \theta) = (1/r) \sum A_\nu \nu(\nu + 1) h_\nu^{(1)}(kr) P_\nu(\cos\theta) \tag{33}$$

**An Application to Antenna Theory.** To place in evidence the usefulness of these complex order wave functions, we shall use them to compute the admittance of a boss antenna. A boss antenna consists of a coaxial line fitted with an infinite flange and a hemispherical boss at the end of the inner conductor (Fig. 1). Except for the addition of a hemispherical boss, it is identical to the circular diffraction antenna.[2]

We assume the antenna is excited by a single propagating mode (the principal mode) in the coaxial region. This mode has no azimuthal variation. The non-vanishing components of the field in the coaxial region, $z \leqq 0$, $a \leqq \rho \leqq b$,

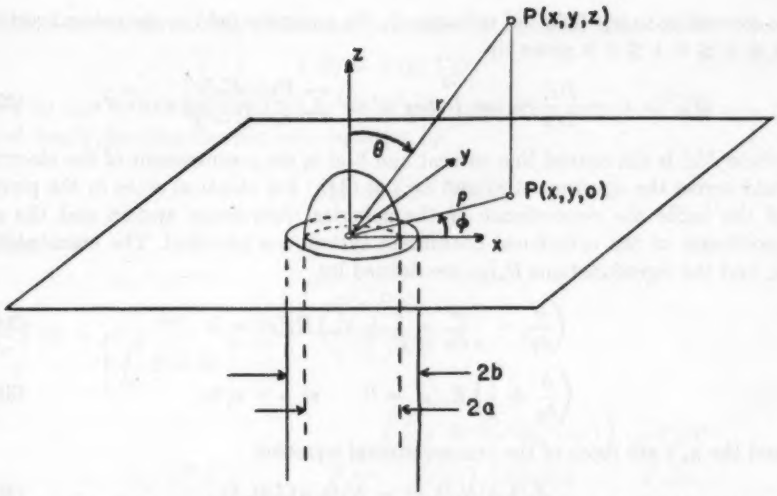



FIG. 1. Boss antenna consisting of coaxial line fitted with flange and hemispherical boss.

are $H_\phi$, $E_\rho$, $E_z$ and in the antenna region, $r \geqq a$, $\theta \leqq \pi/2$, the non-vanishing components are $H_\phi$, $E_r$, $E_\theta$. Cylindrical coordinates are used in the coaxial region and spherical coordinates in the antenna region.

In the antenna region, according to (33), the $r$-component of the electric field is given by

$$E_r(r, \theta) = (1/r)\sum A_\nu \nu(\nu + 1) h_\nu^{(1)}(kr) P_\nu(\cos\theta), \quad r \geqq a, \quad \theta \leqq \tfrac{1}{2}\pi \tag{33a}$$

If we denote $E_r(r, \frac{1}{2}\pi)$ by $\mathcal{E}(r)$, (33a) becomes

$$0 = (1/r)\sum A_\nu \nu(\nu + 1) h_\nu^{(1)}(kr) P_\nu(0), \quad r \geqq b, \quad \theta = \tfrac{1}{2}\pi \tag{34a}$$

$$\mathcal{E}(r) = (1/r)\sum A_\nu \nu(\nu + 1) h_\nu^{(1)}(kr) P_\nu(0), \quad a \leqq r \leqq b, \quad \theta = \tfrac{1}{2}\pi \tag{34b}$$

since the component of the electric field tangent to the perfectly conducting baffle must disappear. Multiplying both sides of (34) by $rh_\nu^{(1)}(kr)$ and integrating from $kr = ka$ to $kr = \infty$, we obtain by virtue of the orthogonality relation, (29),

$$A_\nu = \frac{\int_{ka}^{\infty} r\mathcal{E}(r) h_\nu(kr)\, d(kr)}{\nu(\nu + 1) P_\nu(0) \overline{N_\nu(ka)}} \tag{35}$$

Substituting (35) into (31) we see that the magnetic field at any point in the antenna region is given by

$$H_\phi^{(+)}(r,\theta) = i\omega\epsilon \int_{ka}^{\infty} r'\mathcal{E}(r')\, d(kr') \sum \frac{h_\nu(kr) h_\nu(kr') (\partial/\partial\theta) P_\nu(\cos\theta)}{\nu(\nu + 1) P_\nu(0) N_\nu(ka)} \tag{36}$$

According to eq. (2.34) of reference 2, the magnetic field in the coaxial region $a \leqq \rho \leqq b$, $z \leqq 0$ is given by

$$H_\phi^{(-)}(\rho, z) = \frac{I(z)}{2\pi\rho} + i\omega\epsilon \int_a^b \mathcal{E}(\rho')\rho'\, d\rho' \sum \frac{R_n(\rho)R_n(\rho')}{\sqrt{\lambda_n^2 - k^2}}\, e^{\sqrt{\lambda_n^2 - k^2}\, z} \tag{37}$$

where $I(z)$ is the coaxial line current and $\mathcal{E}(\rho)$ is the $\rho$-component of the electric field across the aperture. $\mathcal{E}(\rho)$ and $\mathcal{E}(r)$ of (34b) are identical since in the plane of the baffle the $r$-coordinate of the spherical coordinate system and the $\rho$-coordinate of the cylindrical coordinate system are identical. The eigenvalues $\lambda_n$ and the eigenfunctions $R_n(\rho)$ are defined by

$$\left(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{1}{\rho^2} + \lambda_n^2\right) R_n(\rho) = 0 \tag{38}$$

$$\left(\frac{\partial}{\partial\rho} + \frac{1}{\rho}\right) R_n(\rho) = 0 \qquad \text{at } \rho = a, b, \tag{39}$$

and the $\lambda_n$'s are roots of the transcendental equation

$$J_0(\lambda_n a)N_0(\lambda_n b) = N_0(\lambda_n a)J_0(\lambda_n b) \tag{40}$$

The orthogonality relations,

$$\int_a^b R_n(\rho)R_m(\rho)\rho\, d\rho = \delta_{nm} \tag{41}$$

$$\int_a^b R_n(\rho)\, d\rho = 0 \tag{42}$$

immediately follow from (38) and (39).

Since $H_\phi$ must be continuous across the aperture, we have $H_\phi^{(+)}(r, \frac{1}{2}\pi) = H_\phi^{(-)}(\rho, 0)$ for $a \leqq \rho \leqq b$. Thus from (36) and (37)

$$\frac{I(0)}{2\pi\rho} + i\omega\epsilon \int_a^b \mathcal{E}(\rho')\rho'\, d\rho' \sum \frac{R_n(\rho)R_n(\rho')}{\sqrt{\lambda_n^2 - k^2}} = i\omega\epsilon k \int_{ka}^\infty r'\mathcal{E}(r')\, dr' \sum \frac{h_\nu(kr)h_\nu(kr')(\partial/\partial\theta)P_\nu(0)}{\nu(\nu+1)P_\nu(0)N_\nu(ka)} \tag{43}$$

This is an integral equation. We shall not attempt to solve it. Rather, we shall use it to formulate a variational principle for the antenna admittance.

From (2.26) and (2.27) of reference 2 the coaxial line current and voltage are given by

$$I(z) = 2\pi(\alpha e^{ikz} + \beta e^{ikz}) \tag{44}$$

$$V(z) = \int_a^b E_\rho^{(-)}(\rho, z)\, d\rho = (\mu/\epsilon)^{\frac{1}{2}}(\alpha e^{ikz} - \beta e^{-ikz}) \log b/a \tag{45}$$

where $\alpha$ and $\beta$ are constants.

The characteristic admittance of the line is given by

$$Y_0 = \frac{2\pi}{(\mu/\epsilon)^{\frac{1}{2}} \log b/a} \tag{46}$$

Now let us derive the variational principle for the admittance $Y(0)$ where

$$Y(0) = I(0)/V(0) \tag{47}$$

We do this by multiplying (43) by $\rho\mathcal{E}(\rho)$ and integrating from $\rho = a$ to $\rho = b$, and finally dividing the resultant equation by

$$\left[\int_a^b \mathcal{E}(\rho)\, d\rho\right]^2 = [V(0)]^2$$

Thus

$$\frac{1}{2\pi} Y(0) + \frac{i\omega\epsilon}{\left[\int_a^b \mathcal{E}(\rho)\, d\rho\right]^2} \sum \frac{\left(\int_a^b \rho\mathcal{E}(\rho) R_n(\rho)\, d\rho\right)^2}{\sqrt{\lambda_n^2 - k^2}} = \frac{i\omega\epsilon k}{\left[\int_a^b \mathcal{E}(\rho)\, d\rho\right]^2} \sum \frac{\left(\int_a^b \rho'\, d\rho' \mathcal{E}(\rho') h_{\nu_n}^{(1)}(k\rho')\right)^2}{\nu(\nu+1) N_\nu(ka)} \frac{\partial}{\partial\theta} P_\nu(0) \over P_\nu(0) \tag{48}$$

It can be easily shown that $Y(0)$ is *stationary* with respect to small variations of $\mathcal{E}(\rho)$ about the true $\mathcal{E}(\rho)$ determined by the integral equation (43).

We choose $1/\rho$ as a trial function. That is, we let $\mathcal{E}(\rho) = 1/\rho$ in (48). Due to the orthogonality relation (42), (48) becomes

$$Y(0) = \frac{2\pi i\omega\epsilon k}{[\log b/a]^2} \sum \frac{\left(\int_a^b h_\nu^{(1)}(k\rho)\, d\rho\right)^2}{\nu(\nu+1) N_\nu(ka)} \frac{(\partial/\partial\theta) P_\nu(0)}{P_\nu(0)} \tag{49}$$

This is an accurate expression for the admittance of the antenna. It is sometimes more useful to consider the ratio $Y(0)/Y_0$, which is a dimensionless quantity.

$$\frac{Y(0)}{Y_0} = \frac{i}{\log (b/a)} \frac{\left(\int_{ka}^{kb} h_\nu^{(1)}(k\rho)\, d(k\rho)\right)^2}{\int_{ka}^{\infty} [h_\nu^{(1)}(k\rho)]^2\, d(k\rho)} \frac{1}{\nu(\nu+1)} \frac{(\partial/\partial\theta) P_\nu(0)}{P_\nu(0)} \tag{50}$$

This is only a formal solution, however. Since the values of the functions involved have not been tabulated, only certain limiting values can be at present obtained for (50). The tabulation of these functions in any detail is quite a worthy project and can best be handled by a high-speed computing group.

## REFERENCES

1. A. Sommerfeld, "Partial Differential Equations", Academic Press, New York, 1949.
2. H. Levine and C. H. Papas, J. App. Phys., **22**, 29-43.

California Institute of Technology
Pasadena, California

(Received September 8, 1953)

# THE ORTHOGONALITY OF THE HYPERBOLOID FUNCTIONS

ROBERT G. VAN NOSTRAND*

**Introduction.** During the course of some recent work in theoretical geophysics, it became necessary to obtain a set of functions which were orthogonal over surfaces in the form of hyperboloids of two sheets. Prolate spheroidal coordinates were employed and an investigation led to a class of associated Legendre function of order $ip - \frac{1}{2}$, of integral degree, and of real argument greater than unity. These functions appeared in a discussion of bowl surfaces and bipolar coordinates by Hobson; and, in a general treatment of zeros, the same author showed that $P^m_{ip-\frac{1}{2}}(\eta)$ has an infinite number of zeros for values of $\eta$ lying between 1 and $\infty$.[1] No other specific reference to these functions has been found. The functions $P^m_{ip-\frac{1}{2}}(\eta)$ can be defined in terms of hypergeometric series or as definite integrals if one adapts the definitions of spherical harmonics of the general type.[1, 2]

In this paper, the functions $P^m_{ip-\frac{1}{2}}(\eta)$ are called hyperboloid functions. A proof of their orthogonality is given; and, this orthogonality property is used to develop an expansion of the reciprocal of the distance between two points in terms of the hyperboloid functions.

**The Property of Orthogonality.** To prove the property of orthogonality, consider two functions of the set and let them be denoted by:

$$H_1 = P^m_{ip-\frac{1}{2}}(\eta), \qquad H_2 = P^m_{iq-\frac{1}{2}}(\eta) \tag{1}$$

Since each of these functions is a solution of the associated Legendre equation, one may write:

$$\begin{aligned} &\frac{d}{d\eta}\left[(1-\eta^2)\frac{dH_1}{d\eta}\right] + \left[(ip-\tfrac{1}{2})(ip+\tfrac{1}{2}) - \frac{m^2}{1-\eta^2}\right]H_1 = 0 \\ &\frac{d}{d\eta}\left[(1-\eta^2)\frac{dH_2}{d\eta}\right] + \left[(iq-\tfrac{1}{2})(iq+\tfrac{1}{2}) - \frac{m^2}{1-\eta^2}\right]H_2 = 0 \end{aligned} \tag{2}$$

Multiplying the first equation by $H_2$, the second by $H_1$, subtracting the first from the second, and integrating from 1 to $\eta$, one gets:

$$\begin{aligned} (p^2 - q^2)\int_1^\eta H_1 H_2 d\eta = &\int_1^\eta H_1 \frac{d}{d\eta}\left[(\eta^2-1)\frac{dH_2}{d\eta}\right]d\eta \\ &- \int_1^\eta H_2 \frac{d}{d\eta}\left[(\eta^2-1)\frac{dH_1}{d\eta}\right]d\eta \end{aligned} \tag{3}$$

If both terms appearing on the right side of the equation are now integrated by parts, equation 3 is easily reduced to:

$$(p^2 - q^2)\int_1^\eta H_1 H_2 d\eta = \left[(\eta^2-1)\left(H_1\frac{dH_2}{d\eta} - H_2\frac{dH_1}{d\eta}\right)\right]_1^\eta \tag{4}$$

* Currently employed as a Senior Research Geophysicist, Magnolia Petroleum Company, Dallas, Texas.

Since $P^m_{ip-\frac{1}{2}}(1)$ and $P'^m_{ip-\frac{1}{2}}(1)$ are finite, the expression on the right is zero at the lower limit. Therefore, the evaluation of the integral on the left is reduced to an investigation of the right side of equation 4 as a function of $\eta$.

Consider now the integral:

$$(5)\quad \lim_{\eta\to\infty}\int_{-\infty}^{\infty} f(q)dq\int_1^{\eta} H_1H_2d\eta = \lim_{\eta\to\infty}\int_{-\infty}^{\infty} f(q)dq\left[\frac{\eta^2-1}{p^2-q^2}\left(H_1\frac{dH_2}{d\eta}-H_2\frac{dH_1}{d\eta}\right)\right]$$

where the function $f(q)$ is assumed to be integrable.

One now requires the asymptotic expansion of the associated Legendre function for large values of the argument. From a general relationship,[2] the expression in this case is seen to be:

$$(6)\quad P^m_{ip-\frac{1}{2}}(\eta) = \{(1/\eta^{\frac{1}{2}})[A_m(p)\cos(p\ln\eta)+B_m(p)\sin(p\ln\eta)]\}\{1+O(1/\eta^2)\}$$

where the symbols introduced are given by:

$$(7)\quad A_m(p) = 2\,Re\left[\frac{2^{ip-\frac{1}{2}}\Gamma(ip)}{\sqrt{\pi}\,\Gamma(ip+\frac{1}{2}-m)}\right],$$

$$B_m(p) = -2\,\mathrm{Im}\left[\frac{2^{ip-\frac{1}{2}}\Gamma(ip)}{\sqrt{\pi}\,\Gamma(ip+\frac{1}{2}-m)}\right]$$

Substituting into equation 5 expressions of the form of equation 6, together with a similar relationship for the derivative of the function with respect to $\eta$, one gets after rearranging terms:

$$\begin{aligned}\lim_{\eta\to\infty}\int_{-\infty}^{\infty} f(q)dq\int_1^{\eta} H_1H_2d\eta &= \lim_{\eta\to\infty}\int_{-\infty}^{\infty} f(q)dq\,\frac{(\eta^2-1)}{\eta^2}\\ &\cdot\left\{\frac{[pA_m(p)B_m(p)-qA_m(q)B_m(p)]}{(p^2-q^2)}\sin(p\ln\eta)\sin(q\ln\eta)\right.\\ &+\frac{[qA_m(p)B_m(q)-pA_m(q)B_m(p)]}{(p^2-q^2)}\cos(p\ln\eta)\cos(q\ln\eta)\\ &+\frac{[A_m(p)A_m(q)-B_m(p)B_m(q)]}{(p+q)}\sin(q\ln\eta)\cos(p\ln\eta)\\ &\left.+\frac{[pA_m(p)A_m(q)+qB_m(p)B_m(q)]}{(p+q)}\,\frac{\sin[(p-q)\ln\eta]}{(p-q)}\right\}\{1+O(1/\eta^2)\}\end{aligned}\tag{8}$$

In evaluating the limit on the right side of the equation, the first point to be noted is that the factor $(\eta^2-1)/\eta^2$ may be removed from under the integral sign and becomes unity in the limit. The second point of interest is that the terms of the order of $1/\eta^2$ vanish in the limit. Therefore, one is left with the sine and cosine terms to be investigated.

It has been shown that the Legendre functions are analytic functions of their order, in this case of $ip-\frac{1}{2}$, within the ranges of their arguments for which the various definitions are valid.[1] It follows then that $A_m(p)$ and $B_m(p)$ are similarly analytic. It is further assumed that the products of $A_m(p)$, $B_m(p)$, and $f(p)$ are integrable.

On this basis then, except in the case for which $q$ equals $p$, it is apparent that the quantity in brackets remains finite for all values of $p$ and $q$. It is now necessary to investigate, term by term, whether an exception exists.

The first term in brackets leads to an indeterminacy, as $q$ approaches $p$, which may be evaluated by l'Hopital's rule:

$$\lim_{q \to p} \frac{pA_m(p)B_m(q) - qA_m(q)B_m(p)}{p^2 - q^2} = \frac{pA_m(p)B'_m(p) - A_m(p)B_m(p) - pA'_m(p)B_m(p)}{-2p} \tag{9}$$

The primes in equation 9 indicate differentiation with respect to $q$. Since $A_m(p)$ and $B_m(p)$ are analytic as explained above, it follows that the first term in brackets remains finite as $q$ approaches $p$. A similar argument may be applied to the second term in the expression; and, the third term requires no discussion of this kind to show that it remains finite. Now, during the integration over $q$, none of these terms adds any contribution, because of a corollary of the Riemann-Lebesgue theorem which states that:[3]

$$\lim_{\mu \to \infty} \int_a^b f(x) \begin{matrix} \sin \\ \cos \end{matrix} (\mu x)\, dx = 0 \tag{10}$$

The effect of the fourth term in brackets is implied by one form of Dirichlet's integrals, which is:[3]

$$\lim_{\mu \to \infty} \int_0^a f(x) \frac{\sin(\mu x)}{x}\, dx = \frac{\pi}{2} f(+0) \tag{11}$$

Therefore, one finds that the net value of the integral in equation 8 is:

$$\lim_{\eta \to \infty} \int_{-\infty}^{\infty} f(q)dq \int_1^{\eta} H_1 H_2 d\eta = \frac{\pi}{2} [A_m^2(p) + B_m^2(p)] f(p) \tag{12}$$

It is now useful to evaluate explicitly the quantity in equation 12 in terms of $m$ and $p$. This end is readily met by noting, from the definitions of $A_m(p)$ and $B_m(p)$, that:

$$A_m^2(p) + B_m^2(p) = \frac{2\Gamma(ip)\Gamma(-ip)}{\pi\Gamma(ip + \frac{1}{2} - m)\Gamma(-ip + \frac{1}{2} - m)} \tag{13}$$

which, through the use of various relationships between the gamma functions,[4] is easily reduced to:

$$A_m^2(p) + B_m^2(p) = \frac{2(-1)^m \Gamma(ip + \frac{1}{2} + m)}{\pi p \Gamma(ip + \frac{1}{2} - m)} \coth(\pi p) \tag{14}$$

Replacing the $H$'s with the corresponding Legendre functions, and substituting equation 14 into equation 12, one finds that:

$$\lim_{\eta \to \infty} \int_{-\infty}^{\infty} f(q)\, dq \int_1^{\eta} P_{ip-\frac{1}{2}}^m(\eta) P_{iq-\frac{1}{2}}^m(\eta)\, d\eta = \frac{(-1)^m \coth(\pi p)}{p} \frac{\Gamma(ip + \frac{1}{2} + m)}{\Gamma(ip + \frac{1}{2} - m)} f(p) \tag{15}$$

As is customary, for example in the case of the orthogonality of the ordinary exponential function, we can now write the orthogonality property of the hyperboloid functions symbolically in terms of the Dirac delta function as follows:

$$(16)\quad \int_1^\infty P^m_{ip-\frac{1}{2}}(\eta)P^m_{iq-\frac{1}{2}}(\eta)\,d\eta = \frac{(-1)^m \coth(\pi p)}{p}\,\frac{\Gamma(ip+\frac{1}{2}+m)}{\Gamma(ip+\frac{1}{2}-m)}\,\delta(p-q)$$

Or, writing out explicitly the factors represented in the ratio of the two gamma functions in equation 16, one may arrive at the alternate form:

$$(17)\quad \int_1^\infty P^m_{ip-\frac{1}{2}}(\eta)P^m_{iq-\frac{1}{2}}(\eta)\,d\eta = \frac{\coth \pi p}{p}\left[p^2+\frac{1}{4}\right]\left[p^2+\frac{9}{4}\right]\cdots\left[p^2+\frac{(2m-1)^2}{4}\right]\cdot\delta(p-q)$$

**Expansion of a Reciprocal Distance.** Using the orthogonality property which has just been derived, it is now possible to expand the reciprocal distance in terms of the hyperboloid functions. To do this, one employs prolate spheroidal coordinates. Varying from 1 to $\infty$, $\eta$ is the ellipsoidal parameter; the parameter in the family of hyperbolas is $\xi$ which varies between plus and minus one; and, $\phi$ which varies from 0 to $2\pi$ is the azimuthal angle.

Let there be a fixed point $P_0$ and a variable point $P$, which are designated by the coordinates $(\eta_0, \xi_0, \phi_0)$ and $(\eta, \xi, \phi)$ respectively, and let the distance between them be designated by $R$. It is known that $1/R$ satisfies Laplace's equation which in prolate spheroidal coordinates is:[5]

$$(18)\quad \frac{\partial}{\partial\eta}\left[(\eta^2-1)\frac{\partial(1/R)}{\partial\eta}\right]+\frac{\partial}{\partial\xi}\left[(1-\xi^2)\frac{\partial(1/R)}{\partial\xi}\right] + \frac{(\eta^2-\xi^2)}{(1-\xi^2)(\eta^2-1)}\frac{\partial^2(1/R)}{\partial\phi^2}=0$$

Other necessary relationships, in this coordinate system, are the $\xi$-component of the gradient:

$$(19)\quad [\nabla(1/R)]_\xi = \frac{1}{b}\sqrt{\frac{1-\xi^2}{\eta^2-\xi^2}}\,\frac{\partial(1/R)}{\partial\xi}$$

and the element of area on a hyperboloidal surface:

$$(20)\quad ds = b^2\sqrt{(1-\xi^2)(\eta^2-\xi^2)}\,d\eta\,d\phi$$

If Laplace's equation be solved for $1/R$, it is found that:

$$(21)\quad \begin{cases} \dfrac{1}{R_1} = \displaystyle\sum_{m=0}^{\infty}\cos m(\phi_0-\phi)\int_{-\infty}^{\infty} C_m(p)P^m_{ip-\frac{1}{2}}(\xi_0)P^m_{ip-\frac{1}{2}}(-\xi)P^m_{ip-\frac{1}{2}}(\eta)\,dp \\ \dfrac{1}{R_2} = \displaystyle\sum_{m=0}^{\infty}\cos m(\phi_0-\phi)\int_{-\infty}^{\infty} C_m(p)P^m_{ip-\frac{1}{2}}(-\xi_0)P^m_{ip-\frac{1}{2}}(\xi)P^m_{ip-\frac{1}{2}}(\eta)\,dp \end{cases}$$

where the subscripts 1 and 2 indicate cases in which $\xi < \xi_0$ and in which $\xi > \xi_0$ respectively. $C_m(p)$ is an arbitrary function of $p$ which also depends on the value

of $m$. The form of these expressions has been so chosen that they are equivalent on the surface of the hyperboloid $\xi = \xi_0$ ; and, the $\xi$-dependence of the expressions has been assigned subject to the behavior of the cone functions $P^m_{ip-\frac{1}{2}}(\xi)$.[5]

$C_m(p)$ can now be evaluated through the use of a sort of boundary condition which may be expressed mathematically in the following equation:

$$\int_S \bar{n}\cdot\nabla(1/R)\, dS = -4\pi \tag{22}$$

where $\bar{n}$ is the unit vector normal outward to the element of surface $dS$ and $\bar{n}\cdot\nabla(1/R)$ is the component of $\nabla(1/R)$ taken normal to that element of area. A second means of expressing the above relationship is to say that $\bar{n}\cdot\nabla(1/R)\, dS$ is the negative of the element of solid angle subtended at the point from which $R$ is measured by the given element of area. The integral is to be taken over any closed surface which encloses the point from which the distance to points on the surface is the variable $R$.

In the present discussion, the central point is $P_0$ while points on the surface, over which the integration is to be performed, are represented by $P$. In order to enclose $P_0$ in this problem, two non-intersecting surfaces are used, one being the hyperboloid $\xi = \xi_0 - \delta$ and the other being the hyperboloid $\xi = \xi_0 + \delta$, plus a segment of an ellipsoid at infinity. Obviously, integration over the part of the enclosing surface at infinity has no effect on the result.

The increment $\delta$ is chosen to be so vanishingly small that it can be neglected in evaluating the various functions over the two remaining surfaces. However, over the first hyperboloid, the expression for $1/R_1$ must be used; while, over the second, the formula for $1/R_2$ is appropriate. For any element of area on one surface, there is a corresponding element, which may be considered to have the same coordinates, on the opposing surface; the only difference is that the normal unit vector is oppositely directed in the two cases. Thus, for any specified coordinates, the total element of solid angle subtended at $P_0$ is given by:

$$\begin{aligned}\{[\nabla(1/R_1)]_\xi - [\nabla(1/R_2)]_\xi\}_{\xi=\xi_0}\, dS = \frac{1}{b}\sqrt{\frac{1-\xi_0^2}{\eta^2-\xi_0^2}}\sum_{m=0}^{\infty}\cos m(\phi_0-\phi) \\ \cdot\int_{-\infty}^{\infty} C_m(p)\{P^m_{ip-\frac{1}{2}}(-\xi_0)P'^m_{ip-\frac{1}{2}}(\xi_0) + P'^m_{ip-\frac{1}{2}}(-\xi_0)P^m_{ip-\frac{1}{2}}(\xi_0)\}P^m_{ip-\frac{1}{2}}(\eta)\, dp\, dS\end{aligned} \tag{23}$$

From a more general relationship,[6] it can be shown that the quantity in brackets on the right side of the equation is given by:

$$\begin{aligned}P^m_{ip-\frac{1}{2}}(-\xi_0)P'^m_{ip-\frac{1}{2}}(\xi_0) + P^m_{ip-\frac{1}{2}}(\xi_0)P'^m_{ip-\frac{1}{2}}(-\xi_0) \\ = -\frac{2(-1)^m\Gamma(ip+\frac{1}{2}+m)\cosh(\pi p)}{\pi(1-\xi_0^2)\Gamma(ip+\frac{1}{2}-m)}\end{aligned} \tag{24}$$

The next step is to multiply both sides of the equation by $P^n_{iq-\frac{1}{2}}(\eta)\cdot\cos n(\phi_0-\phi)$, where $n$ and $q$ are specific numbers in the sets of $m$ and $p$ respectively. Then one should integrate over all values of $\phi$ and $\eta$. First, consider the left side of the equation. The only region in which the quantity in brackets is

appreciably different from zero is in the immediate vicinity of $P_0$—a region so small that the coordinates of all of the points included can be considered to be the same as those of $P_0$. Thus, $P^n_{iq-\frac{1}{2}}(\eta) \cos n(\phi_0 - \phi)$ can be given its value at $P_0$ and can then be removed from under the integral sign as a constant. The remaining integral is then exactly that represented on the left side of equation 22. Therefore, the total result is $-4\pi P^n_{iq-\frac{1}{2}}(\eta_0)$.

Introducing the expression for an element of area and integrating over the angle $\phi$ to eliminate the terms for which $m$ does not equal $n$, one gets for the expression on the right side of equation 23:

$$b\pi(1 + \delta_{0n}) \int_{-\infty}^{\infty} C_n(p) \left[ \frac{-2(-1)^m \Gamma(ip + \frac{1}{2} + m)}{\pi \Gamma(ip + \frac{1}{2} - m)} \cosh(\pi p) \right] dp \cdot \int_1^{\infty} P^n_{ip-\frac{1}{2}}(\eta) P^n_{ip-\frac{1}{2}}(\eta)\, d\eta \tag{25}$$

Employing equation 16, and reverting to the use of $m$ and $p$ instead of $n$ and $q$, one may convert equation 23 into:

$$4\pi P^m_{ip-\frac{1}{2}}(\eta_0) = 2b(1 + \delta_{0m}) C_m(p) \frac{\Gamma^2(ip + \frac{1}{2} + m)}{\Gamma^2(ip + \frac{1}{2} - m)} \frac{\cosh(\pi p) \coth(\pi p)}{p} \tag{26}$$

where the symbol $\Gamma^2$ indicates the square of the gamma function of the argument shown. One can solve immediately for $C_m(p)$ whence it is finally shown that the desired expansion for the reciprocal distance is given by:

$$\frac{1}{R} = \frac{\pi}{b} \sum_{m=0}^{\infty} (2 - \delta_{0m}) \cos m(\phi_0 - \phi) \int_{-\infty}^{\infty} \frac{p \tanh(\pi p)}{\cosh(\pi p)} \cdot \frac{\Gamma^2(ip + \frac{1}{2} - m)}{\Gamma^2(ip + \frac{1}{2} + m)} \begin{Bmatrix} P^m_{ip-\frac{1}{2}}(\xi_0) P^m_{ip-\frac{1}{2}}(-\xi) \\ P^m_{ip-\frac{1}{2}}(-\xi_0) P^m_{ip-\frac{1}{2}}(\xi) \end{Bmatrix} P^m_{ip-\frac{1}{2}}(\eta_0) P^m_{ip-\frac{1}{2}}(\eta)\, dp \tag{27}$$

where the upper term in brackets applies when $\xi < \xi_0$ and the lower when $\xi > \xi_0$.

This expression can be used in any potential function which involves the reciprocal of the distance between two points. The expression is particularly useful in the determination of the arbitrary constants in a potential function due to a point charge or a point source of current in the vicinity of hyperboloidal boundaries between materials of different electrical properties.

The author expresses his gratitude to Dr. Wayne A. Bowers, the Department of Physics, The University of North Carolina, for his help and guidance in the preparation of this paper.

## REFERENCES

1. E. W. Hobson, *Spherical and Ellipsoidal Harmonics* (The University Press, Cambridge, 1931), pp. 178–292, 401–403, and 451–453.
2. Wilhelm Magnus and Fritz Oberhettinger, *Special Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949), pp. 49–75.

3. H. S. Carslaw, *Fourier's Series and Integrals* (Macmillan and Company, London, 1930), pp. 219 and 358.
4. Eugene Jahnke and Fritz Emde, *Tables of Functions* (Dover Publications, New York, 1943), p. 12.
5. William R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, New York, 1950), second edition, pp. 155 and 167–169.
6. Chester Snow, *The Hypergeometric and Legendre Functions* (National Bureau of Standards, 1942), p. 54.

The University of North Carolina at Chapel Hill

(Received December 16, 1952)

# ON THE CALCULATION OF INTEGRALS OF THE FORM

$$\int_0^\theta \sin^p \phi \cos^q \phi \, d\phi$$

By Henry E. Fettis

Integrals of the type $\int_0^\theta \sin^p\phi \cos^q\phi \, d\phi [p > -1, q > -1, \theta \leqq \pi/2]$* occur frequently in physical problems. If $p$ and $q$ are positive integers the integration can be carried out in finite form, while for certain fractional values of "$p$" and "$q$" (e.g. $p = \frac{1}{2}, \frac{1}{3}, q = 0$) the results can be expressed as elliptic integrals. Also if $\theta = \pi/2$, it is well known that the value of this "complete" integral can be written in terms of the Gamma function for any $p$ and $q$ ([1],† formula 855.3). Specifically

$$\int_0^{\pi/2} \sin^p\phi \cos^q\phi \, d\phi = \frac{\Gamma(\frac{1}{2} + \frac{1}{2}p)\Gamma(\frac{1}{2} + \frac{1}{2}q)}{2\Gamma(1 + (\frac{1}{2}p + \frac{1}{2}q))}. \tag{1}$$

However if $\theta \neq \pi/2$, and $p$ and $q$ have arbitrary non integer values, the integral cannot be expressed in finite form in terms of functions previously tabulated.

In this note, the general integral is expressed as a power series in $\sin\theta$ (or $\cos\theta$) which converges when $\theta \leqq \pi/2$, and which may be modified in special cases to obtain more rapid convergence.

Perhaps the simplest way of deriving this series is to write

$$\sin^p\phi \cos^q\phi = \sin^p\phi[1 - \sin^2\phi]^{\frac{1}{2}(q-1)} \cos\phi, \tag{2}$$

expand the second term by the binomial series, and integrate term by term (with $p > -1$) to obtain the desired expansion in powers of $\sin\theta$:

$$\int_0^\theta \sin^p\phi \cos^q\phi \, d\phi$$

$$= \sin^{p+1}\theta\left[\frac{1}{p+1} - \frac{(q-1)}{2(p+3)}\sin^2\theta + \frac{(q-1)(q-3)}{2^2\cdot 2!(p+5)}\sin^4\theta + \cdots\right], \tag{3}$$

since the integrated terms vanish at the lower limit if $p \neq -1$.

Although the above series can be shown to converge for all values of $\theta$, the convergence is poor when $\sin\theta$ is near unity. For values of $\theta$ near $\pi/2$, the change of variable $\phi = \pi/2 - \bar{\phi}$ may be made giving:

$$\int_0^\theta \sin^p\phi \cos^q\phi \, d\phi = \int_{\pi/2-\theta}^{\pi/2} \cos^p\bar{\phi} \sin^q\bar{\phi} \, d\bar{\phi}$$

$$= \int_0^{\pi/2} \sin^q\bar{\phi} \cos^p\bar{\phi} \, d\bar{\phi} - \int_0^{\pi/2-\theta} \sin^q\bar{\phi} \cos^p\bar{\phi} \, d\bar{\phi}. \tag{4}$$

* If $\theta < \pi/2$, the condition $q > -1$ may be dropped.

† Numbers in brackets refer to bibliography at end of paper.

The complete integral may then be expressed in terms of Gamma functions, and the second one calculated from the series (3).

If either $p$ or $q$ is zero, we obtain series expansions for integrals of the type $\int_0^\theta \sin^p\phi\, d\phi$, $\int_0^\theta \cos^p\phi\, d\phi$. Still another series for these latter integrals which converges more rapidly can be found by taking $p = q$ and noting that $\sin^p\phi \cos^p\phi = 2^{-p} \sin^p(2\phi.)$. The variable change $2\phi = \bar{\phi}$, $2\theta = \bar{\theta}$ then gives after some manipulation

$$\int_0^{\bar{\theta}} \sin^p\bar{\phi}\, d\bar{\phi} = (2 \sin \bar{\theta}/2)^{p+1}\left[\frac{1}{p+1} - \frac{p-1}{2^2(p+3)} \sin^2(\bar{\theta}/2) + \frac{(p-1)(p-3)}{2^2 \cdot 2!(p+5)} \sin^4(\bar{\theta}/2) + \cdots + \right]. \tag{5}$$

The above series converges well even when $\theta = \pi/2$, and so may be used with facility to compute values of the integral in the entire interval $0 \leqq \bar{\theta} \leqq \pi/2$.

It may also be noted that the series (3) terminates when $q$ is an odd integer, and since, from (4), the role of $p$ and $q$ may be interchanged, a finite expression can also be obtained if $p$ is an odd integer.

To facilitate the computation of some of these integrals, values of the coefficients $a_n(p)$ which are defined such that

$$\int_0^\theta \sin^p\phi\, d\phi = (2 \sin \theta/2)^{p+1} \sum_{n=0}^{\infty} a_n(p)[\sin(\theta/2)^{2n}] \tag{6}$$

are listed in Table I for $p = -.9(.1).9$, and for such values of $n$ as are necessary to give seven place accuracy in the sum for $\theta < \pi/4$. The coefficients may be checked by means of the relation:

$$\sum_0^\infty a_n(p)(\tfrac{1}{4})^n = \frac{\sqrt{\pi}}{2^{\frac{1}{2}(p+3)}} \frac{\Gamma(\frac{1}{2} + \frac{1}{2}p)}{\Gamma(1 + \frac{1}{2}p)}.$$

As noted earlier, the integral $\int_0^\theta \cos^p\phi\, d\phi$ [or $\int_0^\theta \sin^p\phi\, d\phi$] can be expressed in terms of elliptic integrals for certain special fractional values of $p$. Also, because of well known reduction formulae*, it is sufficient to list these for $-1 < p < 1$. This is done in Table II for these values of $p$ in the interval $-1 < p < 1$ for which the integral is thus expressible.

**Applications**

a. *Projectile Moving in Quasi-Newtonian Medium.* One application of the present integral is in computing the path of a projectile in a medium in which the resistance is proportional to the velocity raised to a given power, $\lambda$. Under this assumption, it can be shown [3] that the velocity, $V$, in terms of the angle of inclination $\theta$ of the projectile to the horizontal is given by

$$V^{-\lambda} = \frac{\cos^\lambda\theta}{\cos^\lambda\theta_0} V_0^{-\lambda} - \frac{C_d\lambda}{mg} \cos^\lambda\theta \int_{\theta_0}^{\theta} \frac{d\phi}{\cos^{1+\lambda}\phi} \tag{7}$$

* See any table of integrals, e.g. Dwight [1] or Pierce [2].

TABLE I

$$\int_0^\theta \sin^p \phi \, d\phi = \left(2 \sin \frac{\theta}{2}\right)^{p+1} \sum_{k=0}^{a_k(p)} a_k(p) \left[\sin\left(\frac{\theta}{2}\right)\right]^{2k}$$

| $k$ | $p=-.9$ | $p=-.8$ | $p=-.7$ | $p=-.6$ | $p=-.5$ | $p=-.4$ | $p=-.3$ | $p=-.2$ | $p=-.1$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 10.0000000 | 5.0000000 | 3.3333333 | 2.5000000 | 2.0000000 | 1.6666667 | 1.4285714 | 1.2500000 | 1.1111111 |
| 1 | .4523810 | .4090909 | .3695652 | .3333333 | .3000000 | .2692308 | .2407407 | .2142857 | .1896552 |
| 2 | .2259147 | .2035714 | .1828488 | .1636363 | .1458333 | .1293478 | .1140957 | .1000000 | .0869898 |
| 3 | .1493136 | .1333064 | .1185615 | .1050000 | .0025480 | .0811363 | .0706996 | .0611765 | .0525091 |
| 4 | .1110405 | .0982728 | .0866177 | .0760000 | .0663488 | .0575973 | .0496827 | .0425455 | .0361295 |
| 5 | .0881618 | .0774236 | .0677048 | .0589292 | .0510254 | .0439261 | .0375685 | .0318934 | .0268452 |
| 6 | .0729764 | .0636524 | .0552785 | .0477702 | .0410754 | .0351060 | .0298058 | .0251160 | .0209819 |
| 7 | .0621778 | .0539060 | .0465284 | .0399658 | .0341452 | .0289985 | .0245059 | .0204807 | .0169978 |
| 8 | .0541135 | .0466603 | .0400540 | .0342146 | .0290687 | .0245483 | .0206273 | .0171404 | .0141433 |
| 9 | .0478667 | .0410713 | .0350819 | .0298180 | .0252060 | .0211784 | .0177048 | .0146362 | .0120147 |
| 10 | .0428883 | .0366346 | .0311512 | .0263568 | .0221782 | .0185486 | .0154344 | .0126997 | .0103757 |
| 11 | .0388297 | .0330313 | .0279707 | .0235671 | .0197476 | .0164460 | .0136267 | .0111614 | .0090821 |
| 12 | .0354590 | .0300489 | .0253477 | .0212748 | .0177577 | .0147312 | .0121581 | .0099219 | .0080394 |
| 13 | .0326159 | .0275416 | .0231499 | .0193606 | .0161018 | .0133092 | .0109446 | .0088990 | .0071841 |
| 14 | .0301863 | .0254055 | .0212833 | .0177401 | .0147045 | .0121132 | .0099274 | .0080444 | .0064720 |
| 15 | .0280866 | .0235649 | .0196797 | .0163520 | .0135113 | .0110907 | .0090641 | .0073214 | .0058715 |

| $k$ | $p=.1$ | $p=.2$ | $p=.3$ | $p=.4$ | $p=.5$ | $p=.6$ | $p=.7$ | $p=.8$ | $p=.9$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | .9090909 | .8333333 | .7692308 | .7142857 | .6666667 | .6250000 | .5882353 | .5555555 | .5263158 |
| 1 | .1451613 | .1250000 | .1060606 | .0882353 | .0714286 | .0555555 | .0405405 | .0263158 | .0128205 |
| 2 | .0639706 | .0538462 | .0445755 | .0361111 | .0284091 | .0214283 | .0151316 | .0094828 | .0044491 |
| 3 | .0375264 | .0311111 | .0253510 | .0202027 | .0156250 | .0115790 | .0080276 | .0049359 | .0022705 |
| 4 | .0252530 | .0206956 | .0166656 | .0131210 | .0100226 | .0073332 | .0050183 | .0030446 | .0013815 |
| 5 | .0184256 | .0149600 | .0119329 | .0093044 | .0070376 | .0050970 | .0034532 | .0020992 | .0009309 |
| 6 | .0141814 | .0114240 | .0090421 | .0069922 | .0052456 | .0037684 | .0025313 | .0015258 | .0006708 |
| 7 | .0113364 | .0090705 | .0071287 | .0054757 | .0040792 | .0029098 | .0019406 | .0011613 | .0005068 |
| 8 | .0093223 | .0074146 | .0057923 | .0044223 | .0032743 | .0023212 | .0015384 | .0009148 | .0003967 |
| 9 | .0078361 | .0061994 | .0048171 | .0036579 | .0026936 | .0018991 | .0012517 | .0007401 | .0003192 |
| 10 | .0067032 | .0052777 | .0040811 | .0030839 | .0022598 | .0015854 | .0010398 | .0006118 | .0002625 |
| 11 | .0058167 | .0045597 | .0035103 | .0026408 | .0019265 | .0013455 | .0008785 | .0005145 | .0002198 |
| 12 | .0051079 | .0039879 | .0030577 | .0022909 | .0016644 | .0011577 | .0007527 | .0004390 | .0001868 |
| 13 | .0045308 | .0035241 | .0026920 | .0020093 | .0014543 | .0010077 | .0006527 | .0003792 | .0001607 |
| 14 | .0040536 | .0031421 | .0023918 | .0017790 | .0012826 | .0008859 | .0005718 | .0003310 | .0001398 |
| 15 | .0036538 | .0028230 | .0021419 | .0015880 | .0011411 | .0007856 | .0005054 | .0002916 | .0001227 |

TABLE II

| $p$ | $\int_0^\theta \cos^p\phi \, d\phi$ | $\psi$ | $k$ |
|---|---|---|---|
| $-\frac{2}{3}$ | $\frac{3^{3/4}}{2} F(\psi, k)$ | $\cos^{-1}\left[\frac{\cos^{2/3}\theta(\sqrt{3}+1)-1}{\cos^{2/3}\theta(\sqrt{3}-1)+1}\right]$ | $\frac{\sqrt{2-\sqrt{3}}}{2} = \sin(\pi/12)$ |
| $-\frac{1}{2}$ | $\sqrt{2}\, F(\psi, k)$ | $\cos^{-1}[\cos^{1/2}\theta]$ | $\frac{\sqrt{2}}{2} = \sin(\pi/4)$ |
| $-\frac{1}{3}$ | $\frac{3^{3/4}}{2} F(\psi, k)$ | $\cos^{-1}\left[\frac{\sqrt{3}-1+\cos^{2/3}\theta}{\sqrt{3}+1-\cos^{2/3}\theta}\right]$ | $\frac{\sqrt{2+\sqrt{3}}}{2} = \sin(5\pi/12)$ |
| $\frac{1}{3}$ | $3^{3/4}\left\{\frac{1-\sqrt{3}}{2} F(\psi, k) + \sqrt{3}\left[E(\psi, k) - \frac{1-\cos\psi}{\sin\psi}\sqrt{1-k^2\sin^2\psi}\right]\right\}$ | $\cos^{-1}\left[\frac{\sqrt{3}-1+\cos^{2/3}\theta}{\sqrt{3}+1-\cos^{2/3}\theta}\right]$ | $\frac{\sqrt{2+\sqrt{3}}}{2} = \sin(5\pi/12)$ |
| $\frac{1}{2}$ | $\sqrt{2}\,[2E(\psi, k) - F(\psi, k)]$ | $\cos^{-1}[\cos^{1/2}\theta]$ | $\frac{\sqrt{2}}{2} = \sin(\pi/4)$ |
| $\frac{2}{3}$ | $\frac{3^{3/4}}{2}\left\{\frac{1+\sqrt{3}}{2} F(\psi. k) - \sqrt{3}\left[E(\psi, k) - \frac{\sin\psi\sqrt{1-k^2\sin^2\psi}}{\cos\psi-(3+\sqrt{3})}\right]\right\}$ | $\cos^{-1}\left[\frac{\cos^{2/3}\theta(\sqrt{3}+1)-1}{\cos^{2/3}\theta(\sqrt{3}-1)+1}\right]$ | $\frac{\sqrt{2-\sqrt{3}}}{2} = \sin(\pi/12)$ |

$$F(\psi, k) = \int_0^\psi \frac{du}{\sqrt{1-k^2\sin^2 u}}$$

$$E(\psi, k) = \int_0^\psi \sqrt{1-k^2\sin^2 u}\, du$$

where $V_0$ is the initial velocity, $g$ the acceleration of gravity, $m$ the mass of the projectile, and $C$ is so defined that

$$\text{Drag} = C_d V^\lambda$$

For a true Newtonian medium, $\lambda = 2$, and the integral in equation (7) can be evaluated in terms of elementary functions. In practical problems however, in which $\lambda$ may actually be empirically determined (and will therefore in all probability not be an integer) the integral may be evaluated by means of the series (3).

b. In the theory of film condensation of vapor at rest on a horizontal cylindric surface [4], there occurs the integral†

$$\int_0^\theta \sin^{1/3}\phi \, d\phi$$

in which $\theta$ is the angle between any radius and the vertical. Although the integral can be expressed in terms of elliptic integrals (Table II), the amplitudes of the elliptic functions are such as to require awkward interpolations if $\theta$ is taken at convenient intervals. It is therefore more satisfactory to evaluate it by means of the series (5). For $p = \frac{1}{3}$, this series becomes

$$\int_0^\theta \sin^{1/3}\phi \, d\phi = \left(2 \sin \frac{\theta}{2}\right)^{4/3}\left[\frac{3}{4} + \frac{1}{2\cdot 5}\sin^2\left(\frac{\theta}{2}\right) + \frac{1\cdot 4}{2\cdot 3\cdot 8\cdot 2!}\sin^4\left(\frac{\theta}{2}\right) + \cdots + \frac{1\cdot 4\cdot 7 \cdots (3n-2)}{2\cdot 3^{n-1} n!(3n+2)}\sin^{2n}\left(\frac{\theta}{2}\right) + \cdots\right]$$

Computations using the above series give values which are slightly larger than these given by Jakob (Loc. cit.), the latter being obtained by graphical integration.

c. *Boundary Layer on Rotating Sphere.* In [5] it is shown that an approximate expression for the boundary layer thickness $\delta$ on a sphere rotating in a viscous fluid is given by

$$\lambda^4 = \frac{231\cdot 5}{\sin^{16/3}\phi \cos^{4/3}\theta}\int_0^\theta \sin^{13/3}\phi \cos^{1/3}\phi \, d\phi$$

where $\lambda = (\Omega/\nu)^{\frac{1}{2}}\delta$, $\nu$ is the kinematic viscosity, $\Omega$ the angular velocity and $\theta$ is a suitably chosen coordinate.

Inserting the series (3) for the integral gives

$$\lambda^4 = \frac{231\cdot 5}{\cos^{4/3}\theta}\left[\frac{3}{16} + \frac{1}{\cdot 22}\sin^2\theta + \frac{1\cdot 4}{3\cdot 2!\cdot 28}\sin^4\theta + \frac{1\cdot 4\cdot 7}{3^2\cdot 3!\cdot 34}\sin^6\theta + \cdots\right]$$

The integral may also be reduced to the one considered in the previous section (b) by two applications of the reduction formula:

$$\int_0^\theta \sin^{13/3}\phi \cos^{1/3}\phi \, d\phi = \frac{-15}{56}\sin^{4/3}\theta \cos^{4/3}\theta - \frac{3}{14}\sin^{10/3}\theta \cos^{4/3}\theta + \frac{5}{14}\left(\frac{1}{2}\right)^{4/3}\int_0^{\theta/2} \sin^{1/3}\phi \, d\phi.$$

† This integral has also been discussed recently by Abramowitz [7].

d. *Application to Statistical Theory.* An integral occurring in the statistical treatment of sampling is the so-called "Student's Integral":

$$I(t, n) = \frac{\Gamma(\frac{1}{2} + \frac{1}{2}n)}{\sqrt{n\pi}\Gamma(\frac{1}{2}\pi)} \int_t^\infty \frac{dx}{[1 + (x^2/n)]^{\frac{1}{2}(n+1)}}$$

(cf. [6], Chapter VI).

By means of the variable change $x^2/n = \cot^2 \phi$, $t^2/n = \cot^2 \theta$ this becomes

$$I(t, n) = \frac{\Gamma(\frac{1}{2} + \frac{1}{2}n)}{\sqrt{\pi}\Gamma(\frac{1}{2}n)} \int_0^\theta \sin^{n-1}\phi \, d\phi$$

Here $n$ represents one less than the number of samples, (often called the number of degrees of freedom) and is therefore an integer. However, if $n$ is large, it is difficult to compute the value of the integral with precision using the conventional formulae in integral tables. Also, it is frequently desirable to know for what value of $\theta$ the integral assumes a pre-assigned small value. For example consider the problem of determining $\theta$ when $n = 9$ and $I(t, n) = .0003$. The equation to be solved is

$$\int_0^\theta \sin^8 \phi \, d\phi = .0003 \sqrt{\pi} \frac{\Gamma(9/2)}{\Gamma(5)} = .0002577$$

Inserting the series (5) for the left side gives

$$\left(2 \sin \frac{\theta}{2}\right)^9 \left[\frac{1}{9} - \frac{7}{22} \sin^2\left(\frac{\theta}{2}\right) + \frac{35}{104} \sin^4\left(\frac{\theta}{2}\right) - \frac{105}{720} \sin^6\left(\frac{\theta}{2}\right) + \frac{105}{6528} \sin^8\left(\frac{\theta}{2}\right) + \cdots\right] = .0002577$$

Since $\theta$ will be small, an approximate value can be found by neglecting all terms in brackets except $\frac{1}{9}$. This gives as a first estimate:

$$(\tfrac{1}{9})(2 \sin \theta/2)^9 = .0002577, \qquad (\sin \theta/2) = .25742$$

Substituting this value into the terms in the brackets gives

$$(\tfrac{1}{9})(2 \sin \theta/2)^9 \times [.823473] = .0002577,$$

$$(\tfrac{1}{9})(2 \sin \theta'/2)^9 = .0003129, \qquad \sin \theta/2 = .26038.$$

Repeating the process gives

$$\sin \theta/2 = .26053,$$

after which no further change in the value is noted. The corresponding value of $t$ is

$$\sqrt{n} \cot \theta = 3 \times 1.718 = 5.154.$$

## REFERENCES

(1) DWIGHT, HERBERT B., *"Tables Of Integral And Other Mathematical Data"*, Macmillan, 1947.

(2) PIERCE, B. O., "*A Short Table Of Integrals*", Ginn and Co., 1929.
(3) POLACHEK, H., "*Solution Of The Differential Equations Of Motion Of A Projectile In A Medium Of Quasi Newtonian Resistance*", Quarterly of Appl. Math., Vol. VII, #3, Oct. 1949.
(4) JAKOB, M., "*Heat Transfer*", John Wiley and Sons, Vol. I, 1949.
(5) HOWARTH, L., "*Note On The Boundary Layer On A Rotating Sphere*", Philosophical Magazine, Series 7, Vol. 42, Pages 1308–15, Nov. 1951.
(6) RIDER, PAUL R., "*An Introduction to Modern Statistical Methods*", J. Wiley & Sons, 1939.
(7) ABRAMOWITZ, M., "*Tables of the Functions* $\int_0^\phi \sin^{1/3}x\,dx$ *and* $\frac{1}{3}\sin^{-4/3}\phi \int_0^\phi \sin^{1/3}x\,dx$". Jour. Res. Natl. Bur. Stand. Vol. 47, pp. 288–290 (Oct. 1951).

WRIGHT AIR DEVELOPMENT CENTER

(Received August 24, 1953; Revised December 30, 1953)

# NEW BOUNDS IN HARMONIC AND BIHARMONIC PROBLEMS*

By L. E. Payne and H. F. Weinberger

**1. Introduction.** The purpose of this paper is to give a method for obtaining arbitrarily close upper and lower bounds for a harmonic function, its gradient, and its Dirichlet integral when the boundary values of the function or its normal derivative are prescribed. These bounds are obtained by means of a harmonic function which approximates the given boundary data.

The principal tools are a Green's identity (3.1), given in another connection by F. Rellich [1], which allows us to find a representation of the Dirichlet integral which differs from that given by the usual Green's theorem; and the simple observation that if $u$ is harmonic, so is $r\partial u/\partial r$. The latter fact permits the evaluation of a boundary integral of the square of the normal derivative of the Green's function, and hence the bounding of a harmonic function. Both of our tools are of such a nature that the boundary must be assumed to be star-shaped with respect to at least one point.

Section 2 treats the simplest example of a Dirichlet integral, the electrostatic capacity in three dimension. The method used here is partly due to Wolf Gross [2]. Some of his numerical results are used in computing upper and lower bounds for the capacity of the cube. The upper bound $(.668a)$ so obtained appears to be the best known to date, (see Polya and Szegö [3], Daboni [4]).

Section 3 gives upper and lower bounds for the Dirichlet integral of a function with (variable) boundary values. Since the best known physical problem of this type is the electrostatic polarization problem [5] we restrict our computations to the exterior problem in three dimensions. However, the method applies equally well to either interior or exterior problems in any number of dimensions.

Section 4 gives a method of finding upper and lower bounds for the Dirichlet integral in the Neumann problem. Again, since the authors are primarily interested in the virtual mass problem [5], the computation is done for the three-dimensional exterior problem. It can, however, also be done in other dimensions and for the interior problem.

There are of course other methods available for obtaining arbitrarily close bounds for the Dirichlet integral. For the Dirichlet problem, an application of the Rayleigh-Ritz procedure to the Dirichlet principle may be used for obtaining upper bounds for the Dirichlet integral. A method for obtaining lower bounds was introduced by Trefftz [6], for estimation of the torsional rigidity. Friedrichs [7] has shown that a lower bound can be derived from the Dirichlet principle by means of a Legendre transformation. Bounds for all the Dirichlet integrals considered in this paper have been obtained in a unified manner from the Schwarz and Bessel inequalities by Diaz and Weinstein [8] and Diaz [9], [10]. A geometric interpretation of these bounds in certain cases has been given by Prager and Synge [11].

* This research was supported by the United States Air Force through the Office of Scientific Research.

In all these methods the upper and lower bounds are obtained separately from two different approximating sequences. The upper bounds involve a complete sequence of functions satisfying boundary conditions on the particular body. Upper bounds involve integrations over the exterior as well as the boundary of the body. The bounds here presented, on the other hand, are both obtained from a single sequence of harmonic functions. The same complete sequence of harmonic functions may be used for all bodies, and the bounds involve integration only on the boundary. We reiterate, however, that, except in the case of capacity, our methods apply only to star-shaped bodies.

In section 5, bounds are developed for the value of the harmonic function itself in terms of its given boundary values or normal derivatives and a harmonic approximating function $\varphi$. Bounds are also given for the gradient of the difference function $u - \varphi$. These involve boundary integrals of the square of the normal derivative or the surface gradient of $u - \varphi$, whichever is given. The boundary must again be star-shaped with respect to some point but not necessarily the point at which the approximation is made. This represents an improvement over the results presented in our technical note [12].

This method is extended in section 6 to the biharmonic clamped plate problem. Upper and lower bounds are given for the deflection of a clamped plate with star-shaped boundary. The bounds here involve a biharmonic function $\varphi$ which is such that its tangential and normal derivatives on the boundary approximate the corresponding given functions for the correct solution $u$ in mean square.

Numerous methods have been proposed for obtaining upper and lower bounds for the value of a harmonic or biharmonic function at a particular point, e.g. Trefftz [13], C. Weber [14], Diaz and Greenberg [15], Greenberg [16], Maple [17], Synge [18], [19], Slobodyanskii [20], [21], Washizu [22] and Fichera [23]. Our method has the advantage that only one sequence of approximating functions is necessary to give both upper and lower bounds. Also all integrations are performed on the boundary, whereas in the other methods the bounds involve area as well as boundary integrals. Our bounds are obviously not as simple as those given by the usual maximum principle for harmonic functions and by the beautiful maximum principle of Miranda [24] for biharmonic functions. On the other hand, the method here presented requires that the boundary data be approximated only in mean square rather than uniformly.

**2. Capacity.** Let $B$ be a closed piecewise differentiable surface in a three-space. Let $r$ be the (Euclidean) distance from an origin interior to $B$. Let $u(x, y, z)$ satisfy

$$\Delta u = 0, \tag{2.1}$$

in the region exterior to $B(\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2)$, and the boundary conditions

$$u = 1 \quad \text{on } B, \tag{2.2}$$

$$u = O(1/r) \quad \text{as} \quad r \to \infty. \tag{2.3}$$

The electrostatic capacity $C$ of $B$ is defined as

$$C = \frac{1}{4\pi} D(u) = -\frac{1}{4\pi} \iint_B u \frac{\partial u}{\partial n}\, ds = -\frac{1}{4\pi} \iint_B \frac{\partial u}{\partial n}\, ds \tag{2.4}$$

where $D(u)$ is the Dirichlet integral over the exterior of $B$ and $n$ the outward normal from $B$. We propose to find arbitrarily close upper and lower bounds for $C$ when $B$ is star-shaped with respect to the origin.

We first prove an identity. Consider the following expansion for $u$

$$u = Cr^{-1} + O(r^{-2}) \quad \text{as} \quad r \to \infty. \tag{2.5}$$

Hence,

$$u + r\partial u/\partial r = O(r^{-2}) \quad \text{as} \quad r \to \infty. \tag{2.6}$$

Furthermore,

$$\Delta(r\partial u/\partial r) = 2\Delta u + r\partial(\Delta u)/\partial r. \tag{2.7}$$

From (2.1) we see that the function $u + r\partial u/\partial r$ is harmonic. If we apply Green's theorem twice and make use of (2.2), (2.3), and (2.6), we find

$$D\left(u, u + r\frac{\partial u}{\partial r}\right) = 0. \tag{2.8}$$

Thus,

$$D(u) = -D\left(u, r\frac{\partial u}{\partial r}\right) = \iint_B r \frac{\partial r}{\partial n}\left(\frac{\partial u}{\partial n}\right)^2 ds \tag{2.9}$$

by (2.2). This can be written by (2.4) as

$$C = \frac{1}{4\pi} \iint_B h\left(\frac{\partial u}{\partial n}\right)^2 ds \tag{2.10}$$

where

$$h = r\partial r/\partial n. \tag{2.11}$$

It is easily seen that $h$ is the distance from the origin to the tangent plane at a particular point of $B$. Thus, the assumption that $B$ is star-shaped with respect to the origin is equivalent to the condition

$$h > 0. \tag{2.12}$$

We proceed to the estimation of $C$ by a method already proposed by Wolf Gross [2] who, however, was unable to obtain any rigorous bounds from it for lack of the identity (2.10). Let $\varphi$ be any function, harmonic in the exterior of $B$ and of $O(r^{-1})$ at $\infty$. If $-(4\pi)^{-1} \iint_B \partial\varphi/\partial n\, ds$ is used to approximate $C$, we find the error

$$C + \frac{1}{4\pi} \iint_B \frac{\partial \varphi}{\partial n}\, ds = \frac{1}{4\pi} \iint_B \frac{\partial u}{\partial n}(\varphi - 1)\, ds \tag{2.13}$$

by Green's theorem.

Then, by Schwarz's inequality and (2.12)

$$\left[C + \frac{1}{4\pi}\iint_B \frac{\partial\varphi}{\partial n}\,ds\right]^2 \leqq \frac{1}{16\pi^2}\iint_B h\left(\frac{\partial u}{\partial n}\right)^2 ds \iint_B h^{-1}(\varphi - 1)^2\,ds. \tag{2.14}$$

Using the identity (2.10)

$$\left[C + \frac{1}{4\pi}\iint_B \frac{\partial\varphi}{\partial n}\,ds\right]^2 \leqq \frac{1}{4\pi} C \iint_B h^{-1}(\varphi - 1)^2\,ds, \tag{2.15}$$

This can be rewritten

$$\left[C + \frac{1}{4\pi}\iint_B \frac{\partial\varphi}{\partial n}\,ds - \frac{1}{8\pi}\iint_B h^{-1}(\varphi - 1)^2\,ds\right]^2 \leqq \frac{1}{64\pi^2}\iint_B h^{-1}(\varphi - 1)^2\,ds\left[\iint_B h^{-1}(\varphi - 1)^2\,ds - 4\iint_B \frac{\partial\varphi}{\partial n}\,ds\right]. \tag{2.16}$$

Equation 2.16 gives upper and lower bounds for $C$. As is more easily seen from (2.14), the bounds can be made arbitrarily close by choosing $\varphi$ so that $\iint_B h^{-1}(\varphi - 1)^2\,ds$ is sufficiently small.

In particular, this can be done by choosing for $\varphi$ a sum of a sufficiently large number $N$ of members of a sequence of harmonic functions $w_\nu$, of $O(r^{-1})$ at $\infty$, whose values on $B$ are complete in $L_2$ norm. Perhaps the most useful such sequence is the sequence of spherical harmonics which is complete on any reasonably regular $B$ [25], [26]. The coefficients of the $w_\nu$ are chosen, by the usual Rayleigh-Ritz technique, to minimize the integral on the right side of (2.14). This minimization results in a function $\varphi$ for which

$$\iint_B h^{-1}(\varphi - 1)w_\nu\,ds = 0, \qquad \nu = 1, \cdots, N. \tag{2.17}$$

Therefore, the equality (2.13) is preserved if any finite sum of $h^{-1}w_1, \cdots, h^{-1}w_n$ is added to $\partial u/\partial n$ on the right hand side. An application of Schwarz's inequality then gives (2.14) with a finite sum of $h^{-1}w_\nu$ added to $\partial u/\partial n$ on the right hand side. We minimize this integral

$$\iint_B h\left(\frac{\partial u}{\partial n} + h^{-1}\sum_{\nu=1}^{N} b_\nu w_\nu\right)^2 ds \tag{2.18}$$

with respect to the coefficients $b_\nu$. The resulting condition is

$$\iint_B h\left(\frac{\partial u}{\partial n} + h^{-1}\sum_{\nu=1}^{N} b_\nu w_\nu\right)(h^{-1}w_\mu)\,ds = 0, \qquad \mu = 1, \cdots, N. \tag{2.19}$$

Since the $w_\nu$ are given functions, the integrals $\iint_B h^{-1}w_\nu w_\mu\,ds$ can be evaluated, while

$$\iint_B w_\mu \frac{\partial u}{\partial n}\,ds = \iint_B \frac{\partial w_\mu}{\partial n}\,ds \tag{2.20}$$

by Green's theorem. Thus, (2.19) is a matrix equation which can be solved for the $b_r$. The resulting minimum of (2.18) is

$$\iint_B \left[ h\left(\frac{\partial u}{\partial n}\right)^2 - h^{-1}\left(\sum_{r=1}^{N} b_r w_r\right)^2 \right] ds = \frac{1}{4\pi} C - \iint_B h^{-1}\left(\sum_{r=1}^{N} b_r w_r\right)^2 ds. \tag{2.21}$$

The inequality (2.16) is then replaced by

$$\begin{aligned} &\left[ C + \frac{1}{4\pi} \iint_B \frac{\partial \varphi}{\partial n}\, ds - \frac{1}{8\pi} \iint_B h^{-1}(\varphi - 1)^2\, ds \right]^2 \\ &\leqq \frac{1}{64\pi^2} \iint_B h^{-1}(\varphi - 1)^2\, ds \iint_B \left[ h^{-1}(\varphi - 1)^2 - 4\frac{\partial \varphi}{\partial n} - 4h^{-1}\left(\sum_{r=1}^{N} b_r w_r\right)^2 \right] ds. \end{aligned} \tag{2.22}$$

This subtraction, which results in considerable numerical improvement of the bounds, was in essence given by W. Gross [2]. We remark that the matrix equation (2.19) involves the same coefficient matrix as that used to find the minimizing function from (2.17), so that little additional work is involved in this improvement. Because of the completeness of the $w_r$, the integral on the left hand side of (2.21) converges to zero, at least if $\partial u/\partial n$ is bounded. This gives an idea of the extent of the improvement involved in the second minimization.

For the well-known problem of the cube, $h$ is a constant, equal to half the edge length. Upper and lower bounds for the cube of edge $a$ can thus be found by using the computations of Gross. They are

$$.627a \leqq C \leqq .668a. \tag{2.23}$$

The lower bound lies below those of $.632a$ given by Polya and Szegö [3], $639a$ obtained by McMahon [27] and $.654a$ given by Daboni [4]. However, the upper bound appears to be the best so far computed.

A particular case of (2.22) in which we take only one function $W$ may be of some computational interest. Putting $\varphi = \alpha w$ and choosing

$$\alpha = \left(\iint_B h^{-1} w\, ds\right)\left(\iint_B h^{-1} w^2\, ds\right)^{-1} \tag{2.24}$$

in order to minimize the integral of $h^{-1}(\varphi - 1)^2$, we obtain

$$\begin{aligned} &\left[ C + \frac{1}{4\pi} \alpha \iint_B \frac{\partial w}{\partial n}\, ds - \frac{1}{8\pi} \iint_B h^{-1}(\alpha w - 1)^2\, ds \right]^2 \\ &\leqq \frac{1}{64\pi^2} \iint_B h^{-1}(\alpha w - 1)^2\, ds \\ &\cdot \left\{ \iint_B \left[ h^{-1}(\alpha w - 1)^2 - 4\alpha \frac{\partial w}{\partial n} \right] ds - 4\left(\iint_B \frac{\partial w}{\partial n}\, ds\right)\left(\iint_B h^{-1} w^2\, ds\right) \right\}. \end{aligned} \tag{2.25}$$

By using the potential function and calculations of Daboni [4] and making only the crude approximation

$$\iint_B h^{-1}(\alpha w - 1)^2\, ds \leqq \text{max. } (\alpha w - 1)^2 \iint_B h^{-1}\, ds \tag{2.26}$$

we obtain the bounds

$$.653a \leqq C \leqq .672a \tag{2.27}$$

for the cube. These bounds compare favorably with the bounds

$$.654a \leqq C \leqq .676a \tag{2.28}$$

obtained by Daboni.

Suppose now that $B$ is not star-shaped with respect to any point, but that it may be divided by interior walls into a finite number of bodies $B_i$ which are star-shaped with respect to points $P_i$. Let the function $h$ at a point on $B$ be defined as the $h$ with respect to $P_i$ on the body $B_i$ on which the point lies. It is easily seen from the maximum modulus theorem that if $u_i$ is the capacity function of $B_i$, then for common points of $B$ and $B_i$, $|\partial u/\partial n| \leqq |\partial u_i/\partial n|$ and hence

$$\iint_B h\left(\frac{\partial u}{\partial n}\right)^2 ds \leqq \sum \iint_{B_i} h\left(\frac{\partial u_i}{\partial n}\right)^2 ds_i = \frac{1}{4\pi}\sum C_i \tag{2.29}$$

where $C_i$ is the capacity of $B_i$. Thus, an upper bound for $\iint_B h(\partial u/\partial n)^2\, ds$ can be found by using some upper bounds for the $C_i$. The result, when inserted in the inequality (2.14), gives upper and lower bounds for $C$ which can be made arbitrarily close by making $\iint_B h^{-1}(\varphi - 1)^2\, ds$ small.

**3. The Dirichlet integral in Dirichlet's problem.** In order to extend the method of section 2 to general boundary values, we need the generalization of the identity (2.10) to this case. The following form of Green's theorem was given by F. Rellich [1] in connection with the membrane problem in $N$ dimensions.

$$\begin{aligned}\int_{R_N} \cdots \int \left[(N-2)\,|\operatorname{grad} u|^2 - \lambda N u^2 - 2r\frac{\partial u}{\partial r}(\Delta u + \lambda u)\right] d\tau_N \\ = \int_{R_{N-1}} \cdots \int \left[2r\frac{\partial u}{\partial r}\frac{\partial u}{\partial n} - r\frac{\partial r}{\partial n}(|\operatorname{grad} u|^2 - \lambda u^2)\right] d\tau_{N-1}\end{aligned} \tag{3.1}$$

where $R_N$ is an $N$-dimensional region bounded by $R_{N-1}$, $n$ the normal on $R_{N-1}$ *into* $R_N$, $u$ is any twice differentiable function in $R_N$ and $\lambda$ any constant. The proof is simply the fact that

$$\begin{aligned}\operatorname{div}\left[\tfrac{1}{2}|\operatorname{grad} u|^2 \operatorname{grad} r^2 - 2r(\partial u/\partial r)\operatorname{grad} u\right] \\ = (N-2)\,|\operatorname{grad} u|^2 - 2r(\partial u/\partial r)\Delta u,\end{aligned} \tag{3.2}$$

$$\operatorname{div}\left[\tfrac{1}{2}u^2 \operatorname{grad} r^2\right] = Nu^2 + 2ur(\partial u/\partial r)$$

and Green's theorem. Following a suggestion of J. B. Diaz, we take $\lambda = 0$ and let

$$\Delta u = 0. \tag{3.3}$$

The identity then becomes

$$(3.4)\quad (N-2)D(u) = \int_{B_{N-1}} \cdots \int \left[ 2r \frac{\partial u}{\partial r}\frac{\partial u}{\partial n} - r\frac{\partial r}{\partial n} \mid \operatorname{grad} u \mid^2 \right] d\tau_{N-1}.$$

Thus, for $N \geqq 3$, (3.4) expresses the Dirichlet integral of $u$ in terms of a boundary integral quite different from the usual representation

$$(3.5)\quad D(u) = -\int_{B_{N-1}} \cdots \int u \frac{\partial u}{\partial n}\, d\tau_{N-1}.$$

Clearly, the integral (3.4), like (3.5), may be used for an exterior domain $R_N$ by taking a limit over spheres at infinity. There is no contribution from infinity, providing $u = 0(r^{-(N-2)})(N \geqq 3)$.

For simplicity of presentation we restrict our attention to the exterior problem for the closed surface $B$ in three dimensions. However, the methods to be presented are easily applied to interior and exterior problems in $N \geqq 2$ dimensions. For example, in two dimensions, the classical problem of the torsional rigidity of a cylindrical beam may be considered [28].

We shall approximate the Dirichlet integral of the three dimensional harmonic function $u$ with assigned boundary values

$$(3.6)\quad u = f \quad \text{on } B, \qquad u = 0(r^{-1}) \quad \text{as} \quad r \to \infty.$$

The identity (3.4) becomes

$$(3.7)\quad \begin{aligned} D(u) &= \iint_B \left[ 2r \frac{\partial u}{\partial r}\frac{\partial u}{\partial n} - r\frac{\partial r}{\partial n} \mid \operatorname{grad} u \mid^2 \right] ds, \\ &= \iint_B \left\{ h\left[ \left(\frac{\partial u}{\partial n}\right)^2 - \mid \operatorname{grad}_t u \mid^2 \right] + 2r \frac{\partial u}{\partial n} \operatorname{grad}_t r \cdot \operatorname{grad}_t u \right\} ds \end{aligned}$$

where $n$ is the outward normal on $B$, $\operatorname{grad}_t$ the tangential projection of the gradient, and $h = r\partial r/\partial n$. It is obvious that for $f =$ constant this reduces to (2.10). We eliminate $D(u)$ between (3.7) and the standard Green's identity

$$(3.8)\quad D(u) = -\iint_B u(\partial u/\partial n)\, ds$$

by subtraction. Completing the square gives

$$(3.9)\quad \begin{aligned} &\iint_B h\left[ \frac{\partial u}{\partial n} + h^{-1}(r \operatorname{grad}_t r \cdot \operatorname{grad}_t f + \tfrac{1}{2}f) \right]^2 ds \\ &\qquad = \iint_B [h \mid \operatorname{grad}_t f \mid^2 + h^{-1}(r \operatorname{grad}_t r \cdot \operatorname{grad}_t f + \tfrac{1}{2}f)^2]\, ds. \end{aligned}$$

We proceed to approximate $D(u)$ by means of a harmonic function of $0(r^{-1})$ at infinity as in Section 2.

$$(3.10)\quad D(u) + \iint_B f \frac{\partial \varphi}{\partial n}\, ds = \iint_B u \frac{\partial(\varphi - u)}{\partial n}\, ds = \iint_B (\varphi - f)\frac{\partial u}{\partial n}\, ds$$

or

$$D(u) + \iint_B f\frac{\partial \varphi}{\partial n}\,ds + \iint_B h^{-1}(\varphi - f)(r \text{ grad}, r \cdot \text{grad}, f + \tfrac{1}{2}f)\,ds$$

(3.11)

$$= \iint_B (\varphi - f)\left[\frac{\partial u}{\partial n} + h^{-1}(r \text{ grad}, r \cdot \text{grad}, f + \tfrac{1}{2}f)\right] ds.$$

Applying Schwarz's inequality and using (3.9) we find

$$\left\{D(u) + \iint_B \left[f\frac{\partial \varphi}{\partial n} + h^{-1}(\varphi - f)(r \text{ grad}, r \cdot \text{grad}, f + \tfrac{1}{2}f)\right] ds\right\}^2$$

(3.12)

$$\leqq \iint_B h^{-1}(\varphi - f)^2\,ds \iint_B [h\,|\,\text{grad}, f\,|^2 + h^{-1}(r \text{ grad}, r \cdot \text{grad}, f + \tfrac{1}{2}f)^2]\,ds.$$

The second integral on the right can be computed from the given boundary values $f$. To simplify this computation it should be noted that if $f$ is known as a continuously differentiable function in a neighborhood of $B$ as well as on $B$, then

$$|\text{ grad}, f\,|^2 = |\text{ grad } f\,|^2 - (\partial f/\partial n)^2$$

(3.13)

$$r \text{ grad}, r \cdot \text{ grad}, f = r(\partial f/\partial r) - h(\partial f/\partial n).$$

The first term on the right of (3.12) can be made arbitrarily small by a suitable choice of $\varphi$.

The Rayleigh-Ritz technique for minimizing $\iint_B h^{-1}(\varphi - f)^2\,ds$ consists of letting $\varphi = \sum_{\nu=1}^{N} d_\nu w_\nu$ where $\{w_\nu\}$ is a sequence of regular harmonic functions complete on $B$. The minimization condition is

$$\iint_B h^{-1}(\varphi - f)w_\mu\,ds = 0, \qquad \mu = 1, \cdots, N. \tag{3.14}$$

For this reason, the equation (3.11) remains valid if a linear combination $h^{-1}\sum_{\nu=1}^{N} d_\nu w_\nu$ is added to the bracketed term on the right hand side. When Schwarz's inequality is applied, the right-hand side can be minimized with respect to the coefficients $d_\nu$. The result is that (3.12) is replaced by the better inequality,

$$\left\{D(u) + \iint_B \left[f\frac{\partial \varphi}{\partial n} + h^{-1}(\varphi - f)(r \text{ grad}, r \cdot \text{grad}, f + \tfrac{1}{2}f)\right] ds\right\}^2$$

$$\leqq \iint_B h^{-1}(\varphi - f)^2\,ds$$

(3.15)

$$\cdot \left\{\iint_B [h\,|\,\text{grad}, f\,|^2 + h^{-1}(r \text{ grad}, r \cdot \text{grad}, f + \tfrac{1}{2}f)^2]\,ds\right.$$

$$\left. - \iint_B h^{-1}\left(\sum_{\nu=1}^{N} d_\nu w_\nu\right)^2 ds\right\},$$

where $d_\nu$ are the minimizing coefficients.

**4. The Dirichlet integral in Neumann's Problem.** We now consider the boundary value problem

$$\begin{aligned} &\Delta u = 0 \qquad \text{exterior to } B, \\ &\partial u/\partial n = g \qquad \text{on } B, \qquad u = 0(r^{-2}) \text{ as } r \to \infty, \end{aligned} \tag{4.1}$$

where $B$ is again a star-shaped closed surface in three-space as in Section 2. We again eliminate $D(u)$ by subtracting (3.8) from (3.7) and apply Schwarz's inequality.

$$\begin{aligned} &\left| \iint_B h \left[ |\operatorname{grad}_s u|^2 - \left(\frac{\partial u}{\partial n}\right)^2 \right] ds \right| \\ &\leqq \left\{ 4 \iint_B h |\operatorname{grad}_s u|^2 \, ds \iint_B h^{-1} r^2 \left(\frac{\partial u}{\partial n}\right)^2 |\operatorname{grad}_s r|^2 \, ds \right\}^{\frac{1}{2}} \\ &\quad + \left\{ \iint_B \psi u^2 \, ds \iint_B \psi^{-1} \left(\frac{\partial u}{\partial n}\right)^2 ds \right\}^{\frac{1}{2}}, \end{aligned} \tag{4.2}$$

where $\psi$ is any positive function.

Since $B$ is star-shaped, one can use the angle variables of a spherical coordinate system as coordinates on $B$. Then all integrals can be considered to be on the unit sphere. We choose $\psi$ so that

$$\psi \, ds = d\omega, \tag{4.3}$$

where $d\omega$ is the projection of the element of area on the unit sphere. This entails

$$\psi = hr^{-3}. \tag{4.4}$$

The Dirichlet integral is unchanged if a constant is added to $u$, and, since $u = 0(r^{-2})$, the form (3.8) still holds. We choose this constant so that

$$\iint u \, d\omega = 0. \tag{4.5}$$

It is well known that the minimum problem on the unit sphere

$$\min \frac{\iint |\overline{\operatorname{grad}}\, u|^2 \, d\omega}{\iint u^2 \, d\omega}, \tag{4.6}$$

where $\overline{\operatorname{grad}}$ is the surface gradient on the sphere, has eigenvalues $n(n+1)$ with surface harmonics as eigenfunctions [29, p. 272]. The condition (4.5) requires $u$ to be orthogonal to the constant, which corresponds to the eigenvalue zero. Hence,

$$\iint u^2 \, d\omega \leqq \tfrac{1}{2} \iint |\overline{\operatorname{grad}}\, u|^2 \, d\omega. \tag{4.7}$$

[1] This implies $\iint_B g \, ds = 0$. This can always be attained by subtracting $(4\pi r)^{-1} \iint_B g \, ds$ from the original unknown function. The change of Dirichlet integral can be computed explicitly.

Elementary geometric considerations show that

$$|\operatorname{grad}_s u| \geqq r^{-1} |\operatorname{grad} u|, \tag{4.8}$$

and hence

$$\begin{aligned} \iint_B \psi u^2 \, dS = \iint u^2 \, dw &\leqq \tfrac{1}{2} \iint_B hr^{-1} |\operatorname{grad}_s u|^2 \, dS \\ &\leqq \tfrac{1}{2} \max (r^{-1}) \iint_B h |\operatorname{grad}_s|^2 \, dS. \end{aligned} \tag{4.9}$$

This inequality when substituted in (4.2) gives a quadratic inequality for $\iint_B h |\operatorname{grad}_s u|^2 \, dS$. Substituting this back into (4.9) gives an upper bound for $\iint_B \psi u^2 \, dS$ involving only integrals of the known function $g$ and the geometry of $B$.

We now approximate $D(u)$ by $-\iint_B \varphi g \, dS$ where $\varphi$ is a regular harmonic function. Then

$$D(u) + \iint_B \varphi g \, dS = \iint_B (\varphi - u) \frac{\partial u}{\partial n} \, dS = \iint_B u \frac{\partial}{\partial n} (\varphi - u) \, dS, \tag{4.10}$$

and

$$\left\{ D(u) + \iint_B \varphi g \, dS \right\}^2 \leqq \iint_B \psi^{-1} \left( \frac{\partial \varphi}{\partial n} - g \right)^2 dS \iint_B \psi u^2 \, dS. \tag{4.11}$$

The second integral on the right has been explicitly bounded. The first is to be made small by choice of the function $\varphi$ as before.

The Rayleigh-Ritz technique and the improvement of the bound by the addition of arbitrary functions to $\varphi$ in (4.10) can be carried through as in Sections 2 and 3.

**5. Bounds for a harmonic function.** In a space of $N \geqq 2$ dimensions let $B$ be a piecewise smooth $(N-1)$-dimensional boundary which is star-shaped with respect to the origin 0. We are given the Dirichlet problem

$$\Delta u = 0 \quad \text{inside } B, \tag{5.1}$$

$$u = f \quad \text{on } B. \tag{5.2}$$

(The exterior problem is handled in the same way.)

The function $u$ is to be estimated by means of a harmonic function $\varphi$ whose boundary values approximate $f$ in mean square. We shall give a bound for $|u - \varphi|$ at any point $P$ interior to $B$.

Let $r$ be the distance from 0 and again put $h = r\partial r/\partial n$ on $B$ so that $h > 0$. The Green's function for the region inside $B$ has the form

$$G(P, Q) = \begin{cases} (2\pi)^{-1} \ln r_{PQ} + g(P, Q) & \text{for } N = 2 \\ -(\omega_N r_{PQ}^{N-2})^{-1} + g(P, Q) & \text{for } N \geqq 3 \end{cases} \tag{5.3}$$

where $\omega_N$ denotes the surface of the unit sphere in $N$ dimensions, $r_{PQ}$ is the distance from $P$ to $Q$ and $g(P, Q)$ is a regular harmonic function of both $P$ and $Q$. Thus we find

$$r_{OP} \frac{\partial G(P, Q)}{\partial r_{OP}} + r_{OQ} \frac{\partial G(P, Q)}{\partial r_{OQ}} + (N - 2)G(P, Q)$$

$$= \begin{cases} \dfrac{1}{2\pi} + r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}} + r_{OQ} \dfrac{\partial g(P, Q)}{\partial r_{OQ}}, & N = 2 \\ r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}} + r_{OQ} \dfrac{\partial g(P, Q)}{\partial r_{OQ}} + (N - 2)g(P, Q), & N \geqq 3 \end{cases} \tag{5.4}$$

which is a regular harmonic function of $P$ (see (2.7)). Thus by Green's theorem

$$\iint_B \left[ r_{OP} \frac{\partial G(P, Q)}{\partial r_{OP}} + r_{OQ} \frac{\partial G(P, Q)}{\partial r_{OQ}} + (N - 2)G(P, Q) \right] \frac{\partial G(P, Q)}{\partial n} \, d\tau_{N-1}(Q)$$

$$= \begin{cases} \dfrac{1}{2\pi} + 2r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}}\Big|_{Q \to P} & N = 2 \\ (N - 2)g(P, P) + 2r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}}\Big|_{Q \to P} & N \geqq 3. \end{cases} \tag{5.5}$$

Since by definition $G(P, Q)$ vanishes for $Q$ on $B$, (5.5) may be rewritten

$$\iint_B h \left( \frac{\partial G(P, Q)}{\partial n} \right)^2 d\tau_{N-1}(Q) \begin{cases} \dfrac{1}{2\pi} + 2r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}}\Big|_{Q \to P} & N = 2 \\ (N - 2)g(P, P) + 2r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}}\Big|_{Q \to P} & N \geqq 3. \end{cases} \tag{5.6}$$

If $\psi$ is any harmonic function, then by Green's theorem

$$\psi(P) = \iint_B \psi \frac{\partial G}{\partial n} \, d\tau_{N-1}. \tag{5.7}$$

By Schwarz's inequality and (5.6),

$$|\psi(P)|^2 \leqq \iint_B h^{-1} \psi^2 \, d\tau_{N-1} \begin{cases} \left[ \dfrac{1}{2\pi} + 2r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}}\Big|_{Q \to P} \right] & \text{for } N = 2 \\ \left[ (N - 2)g(P, P) + 2r_{OP} \dfrac{\partial g(P, Q)}{\partial r_{OP}}\Big|_{Q \to P} \right] & \text{for } N \geqq 3 \end{cases} \tag{5.8}$$

To estimate the second term on the right, which is the value at $P$ of a harmonic function with known boundary values, let $\psi$ be this harmonic function. We find

$$\frac{1}{2\pi} + 2r_{OP}\frac{\partial g(P,Q)}{\partial r_{OP}}\bigg|_{Q=P} \leqq \oint_B h^{-1}\left[\frac{1}{2\pi} + 2r_{OP}\frac{\partial}{\partial r_{OP}}\left(-\frac{1}{2\pi}\ln r_{PQ}\right)\right]^2 d\tau_1(Q), \quad \text{for } N = 2$$

$$(5.9)\quad (N-2)g(P,P) + 2r_{OP}\frac{\partial g(P,Q)}{\partial r_{OP}}\bigg|_{Q=P} \leqq \iint_B h^{-1}\left[(N-2)(\omega_N r_{PQ}^{N-2})^{-1} + 2r_{OP}\frac{\partial}{\partial r_{OP}}(\omega_N r_{PQ}^{N-2})^{-1}\right]^2 d\tau_{N-1}(Q) \quad \text{for } N \geqq 3.$$

Letting $\psi = u - \varphi$ in (5.8), using (5.9) and carrying out the differentiations yields

$$(5.10)\quad |u(P) - \varphi(P)|^2 \leqq \iint_B h^{-1}(f-\varphi)^2\, d\tau_{N-1} \begin{cases} \oint_B (4\pi^2 h^{-1}[r_{PQ}^{-2}(r_{OQ}^2 - r_{OP}^2)]^2\, d\tau_1(Q), & N = 2 \\ \iint_B (N-2)^2(\omega_N^2 h)^{-1}[r_{PQ}^{-N}(r_{OQ}^2 - r_{OP}^2)]^2\, d\tau_{N-1}(Q), & N \geqq 3. \end{cases}$$

This gives upper and lower bounds for $u(P)$. The first integral on the right is, of course, to be minimized. If Rayleigh-Ritz technique is used, then, as in preceding sections, a linear combination $h^{-1}\sum d_s w_s$ may be subtracted from $(\partial G/\partial n)$ in (5.7) without disturbing the equality. The integral $\iint_B h[(\partial G/\partial n) - h^{-1}\sum d_s w_s]^2\, d\tau_{N-1}$ can then be minimized (Note that $\iint_B w_s(\partial G/\partial n)\, d\tau_{N-1} = w_s(P)$). The integral $\iint h^{-1}(\sum d_s w_s)^2\, d\tau_{N-1}$ is subtracted from the second factor on the right hand side of (5.10), which may give considerable improvement of the bounds.

Since the derivatives $(\partial/\partial x_i)(u-\varphi)$ are also harmonic, we may insert them for $u - \varphi$ in (5.10). Summing with respect to $i$ gives an inequality for $|\text{ grad }(u-\varphi)|^2$ in terms of $\iint_B h^{-1}|\text{ grad }(u-\varphi)|^2\, d\tau_{N-1}$. We thus need a bound for this integral. Since $|\text{ grad }(u-\varphi)|^2 = |\text{ grad}_s(u-\varphi)|^2 + [\partial(u-\varphi)/\partial n]^2$, the former of which is known, it is only necessary to bound $\iint_B h^{-1}[\partial(u-\varphi)/\partial n]^2\, d\tau_{N-1}$. Such a bound is obtained by subtracting $(N-2)$ times (3.5) from (3.4) with $u$ replaced by $u - \varphi$, decomposing $\partial(u-\varphi)/\partial r$, and applying Schwarz's inequality.

$$\iint_B h\left[\frac{\partial}{\partial n}(u-\varphi)\right]^2 d\tau_{N-1} \leqq \iint_B \{h\,|\,\text{grad}_*\,(u-\varphi)\,|^2$$

$$+ \tfrac{1}{2}h^{-1}[2r\,\text{grad}_*\,r\cdot\text{grad}_*\,(u-\varphi)+(N-2)(u-\varphi)]^2\}\,d\tau_{N-1}$$

$$+\frac{1}{2}\left\{\iint_B h^{-1}[2r\,\text{grad}_*\,r\cdot\text{grad}_*\,(u-\varphi)+(N-2)(u-\varphi)]^2\,d\tau_{N-1}\right\}^{\frac{1}{2}} \tag{5.11}$$

$$\times\left\{\iint_B (h^{-1}[2r\,\text{grad}_*\,r\cdot\text{grad}_*\,(u-\varphi)+(N-2)(u-\varphi)]^2 + 4h\,|\,\text{grad}_*\,(u-\varphi)\,|^2)\,d\tau_{N-1}\right\}^{\frac{1}{2}}.$$

Since $h$ is bounded below, we can say

$$\iint_B h^{-1}\left[\frac{\partial}{\partial n}(u-\varphi)\right]^2 d\tau_{N-1} \leqq \max(h^{-2})\iint_B h\left[\frac{\partial}{\partial n}(u-\varphi)\right]^2 d\tau_{N-1}. \tag{5.12}$$

Finally,

$$|\,\text{grad}\,(u(P)-\varphi(P))\,|^2$$

$$\leqq\left[\iint_B h^{-1}\,|\,\text{grad}_*\,(u-\varphi)\,|^2\,d\tau_{N-1}+\max(h^{-2})\right.$$

$$\left.\cdot\iint_B h\left[\frac{\partial}{\partial n}(u-\varphi)\right]^2 d\tau_{N-1}\right]\times\left[\iint_B h^{-1}r_{PQ}^{-2N}(r_{OQ}^2-r_{OP}^2)^2\,d\tau_{N-1}(Q)\right] \tag{5.13}$$

$$\times\begin{cases}(4\pi^2)^{-1} & \text{for } N=2\\ (N-2)^2w_N^{-2} & \text{for } N\geqq 3\end{cases}$$

which together with (5.11) gives a bound that can be made arbitrarily small by making $|\,\text{grad}_*\,(u-\varphi)\,|^2$ small in mean.

These considerations give, for example, a method for approximating stresses in the torsion problem.

In the case of the Neumann problem, inequalities of the form (4.2) and (4.9) are easily derived for $N$ dimensions. (In 2 dimensions, the analogue of (4.9) is, of course, simpler.) Thus, $\iint_B h^{-1}(u-\varphi)^2\,d\tau_{N-1}$ can be bounded in terms of integrals involving $[\partial(u-\varphi)/\partial n]^2$ and the geometry of $B$. Thus, a bound for $|\,u-\varphi\,|$ is found from (5.10) which can be made arbitrarily small by choosing $\partial\varphi/\partial n$ to approximate $\partial u/\partial n$ in mean square. The first derivatives can also be approximated with known error.

**6. Bounds for a biharmonic function.** We consider the classical static clamped plate problem in two dimensions, i.e.

$$\Delta\Delta u = 0 \qquad \text{inside } B, \tag{6.1}$$

$$u = f, \qquad \partial u/\partial n = g \qquad \text{on } B. \tag{6.2}$$

(The non-homogeneous problem $\Delta\Delta v = F$, $v = f$, $\partial v/\partial n = g$ on $B$ may be reduced to this form by considering the function $u = v - (8\pi)^{-1}\iint Fr^2 \ln r \, ds$). We shall find arbitrarily close bounds for $u$ at any point $P$ inside $B$ provided $B$ is star-shaped with respect to an interior point 0 which is chosen as the origin. These bounds will be expressed in terms of $f$, $g$, and an arbitrary biharmonic function $\varphi$. Let

$$\psi = u - \varphi \tag{6.3}$$

and let $w$ be the harmonic function which takes the value of $\psi$ on $B$ so that

$$\Delta\Delta(\psi - w) = 0 \text{ in } B, \qquad \psi - w = 0 \text{ on } B \tag{6.4}$$

Then by Green's theorem

$$\psi(P) - w(P) = -\oint_B \frac{\partial}{\partial n}(\psi - w)\Delta\Gamma \, ds, \tag{6.5}$$

where $\Gamma(P, Q)$ is the biharmonic Green's function defined by

$$\begin{aligned} \Gamma(P, Q) &= (8\pi)^{-1} r_{PQ}^2 \ln r_{PQ} + \gamma(P, Q) \qquad \text{inside } B \\ \Gamma &= \partial\Gamma/\partial n = 0 \qquad \text{on } B, \end{aligned} \tag{6.6}$$

the quantity $\gamma(P, Q)$ being a regular biharmonic function in $B$.

$$\Gamma(P, Q) - \tfrac{1}{2} r_{OP} \frac{\partial\Gamma(P, Q)}{\partial r_{OP}} - \tfrac{1}{2} r_{OQ} \frac{\partial\Gamma(P, Q)}{\partial r_{OQ}}$$

is a regular biharmonic function taking the value $\gamma(P, P) - r_{OP}\partial\gamma(P, Q)/\partial r_{OP}\,|_{Q=P}$ at $P$ and vanishing on $B$. Furthermore, the normal derivative on $B$ is given by

$$-\tfrac{1}{2} h \Delta\Gamma \tag{6.7}$$

Then by Green's theorem

$$\gamma(P, P) - r_{OP} \frac{\partial\gamma(P, Q)}{\partial r_{OP}}\bigg|_{Q=P} = \tfrac{1}{2} \oint_B h[\Delta\Gamma(P, Q)]^2 \, ds_Q. \tag{6.8}$$

If we apply Schwarz's inequality to (6.5) and substitute (6.8) we obtain

$$\begin{aligned} |\psi(P) - w(P)|^2 \leqq 2 \oint_B h^{-1} \left[\frac{\partial}{\partial n}(\psi - w)\right]^2 ds \\ \left\{\gamma(P, P) - r_{OP} \frac{\partial\gamma(P, Q)}{\partial r_{OP}}\bigg|_{Q=P}\right\}. \end{aligned} \tag{6.9}$$

Now by (5.11) we have, since $w$ is harmonic and takes the boundary value $\psi$,

$$\begin{aligned} \oint_B h\left(\frac{\partial w}{\partial n}\right)^2 ds \leqq \oint_B \left[h\left(\frac{\partial\psi}{\partial s}\right)^2 + 2h^{-1}\left(r\frac{\partial r}{\partial s}\frac{\partial\psi}{\partial s}\right)^2\right] ds \\ + \left\{\oint_B h^{-1}\left(r\frac{\partial r}{\partial s}\frac{\partial\psi}{\partial s}\right)^2 ds\right\}^{\frac{1}{2}} \left\{\oint_B \left[4h^{-1}\left(r\frac{\partial r}{\partial s}\frac{\partial\psi}{\partial s}\right)^2 + 4h\left(\frac{\partial\psi}{\partial s}\right)^2\right] ds\right\}^{\frac{1}{2}} \end{aligned} \tag{6.10}$$

Clearly

$$\oint_B h^{-1}\left(\frac{\partial w}{\partial n}\right)^2 ds \leqq \max\,(h^{-2}) \oint_B h\left(\frac{\partial w}{\partial n}\right)^2 ds. \tag{6.11}$$

By the triangle inequality

$$\begin{aligned}\left\{\oint_B h^{-1}\left[\frac{\partial}{\partial n}(\psi - w)\right]^2 ds\right\}^{\frac{1}{2}} &\leqq \left\{\oint_B h^{-1}\left(\frac{\partial \psi}{\partial n}\right)^2 ds\right\}^{\frac{1}{2}} + \left\{\oint_B h^{-1}\left(\frac{\partial w}{\partial n}\right)^2 ds\right\}^{\frac{1}{2}} \\ &\leqq \left\{\oint_B h^{-1}\left(\frac{\partial \psi}{\partial n}\right)^2 ds\right\}^{\frac{1}{2}} + \max\,(h^{-1})\left\{\oint_B h\left(\frac{\partial w}{\partial n}\right)^2\right\}^{\frac{1}{2}}.\end{aligned} \tag{6.12}$$

Furthermore, by (5.10),

$$w^2(P) \leqq \frac{1}{4\pi^2}\oint_B h^{-1}\psi^2\, ds \oint_B h^{-1} r_{PQ}^{-4}(r_{OQ}^2 - r_{OP}^2)^2\, ds_Q, \tag{6.13}$$

and by the triangle inequality

$$|\psi(P)| \leqq |\psi(P) - w(P)| + |w(P)|. \tag{6.14}$$

Substitution of (6.9), (6.10), (6.11) and (6.13) in (6.14) gives a bound for $|\psi(P)|$ involving boundary integrals of the squares of $\psi$, $\partial\psi/\partial n$, and $\partial\psi/\partial s$, together with the unknown quantity $\gamma(P, P) - r_{OP}\partial\gamma(P, Q)/\partial r_{OP}$. A bound for the latter quantity may be obtained by using (6.9) with

$$u(Q) = \Gamma(P, Q) - r_{OP}\frac{\partial}{\partial r_{OP}}\Gamma(P, Q) - \frac{1}{4}\left[r_{PQ}^2 - r_{OP}\frac{\partial}{\partial r_{OP}}(r_{PQ}^2)\right] G(P, Q). \tag{6.15}$$

In this equation $G(P, Q)$ is the harmonic Green's function; hence $u$ is regular, biharmonic, and vanishes on $B$. Then we have

$$\gamma(P, P) - r_{OP}\frac{\partial\gamma(P, Q)}{\partial r_{OP}}\bigg|_{Q=P} \leqq \frac{1}{8}\oint_B h^{-1}\left(\frac{\partial G}{\partial n}\right)^2 (r_{OQ}^2 - r_{OP}^2)^2\, ds_Q. \tag{6.16}$$

From (5.6) and (5.9)

$$\begin{aligned}\gamma(P, P) - r_{OP}\frac{\partial\gamma(P, Q)}{\partial r_{OP}}\bigg|_{Q=P} &\leqq \frac{1}{32\pi^2}\max_{Q\in B}(h^{-2}(r_{OQ}^2 - r_{OP}^2)^2) \\ &\quad \times \oint_B h^{-1} r_{PQ}^{-4}(r_{OQ}^2 - r_{OP}^2)\, ds_Q.\end{aligned} \tag{6.17}$$

Thus, we have found an explicit bound for $|u - \varphi|$ in terms of the given values $f$ and $g$ and the arbitrarily chosen biharmonic function $\varphi$. Furthermore, this bound can be made arbitrarily small by choosing $\varphi$ in such a way that it approximates $f$, $\partial\varphi/\partial s$ approximates $\partial f/\partial s$, and $\partial\varphi/\partial n$ approximates $g$ in mean square [24, 30, 31, 32].

The extension of these methods to higher dimensions and to solutions of polyharmonic equations is immediate. It appears likely that they can be also extended to systems of differential equations, such as those arising in elasticity.

## BIBLIOGRAPHY


1. Rellich, F.: *Darstellung der Eigenwerte von* $\Delta u + \lambda u$ *durch ein Rand integral*, Math. Zeit. **46** (1940), pp. 635–646.
2. Gross, W.: *Sul calcolo della capacita elettrostatica di un conduttore*, Rend. Acc. Naz. Lincei **12** (1952), pp. 496–506.
3. Polya, G. and Szegö, G.: *Isoperimetric inequalities in mathematical physics*, Annals of Mathematics Studies Number 27, Princeton University Press (1951).
4. Daboni, L.: *Applicazione al caso del cubo di un metodo per il calcolo per eccesso e per difetto della capacita elettrostatica di un conduttore*, Atti. Accad. Nazi. Lincei Rend. Cl. Sci Fis. Mat. Nat. (8) **14** (1953), p. 461–466.
5. Schiffer, M. and Szegö, G.: *Virtual mass and polarization*, Trans. Amer. Math. Soc., **67** (1949), pp. 130–205.
6. Trefftz, E.: *Ein Gegenstück zum Ritzschen Verfahren*, Proc. Second Int. Congress Appl. Mech., Zürich, (1927), pp. 131–137.
7. Friedrichs, K. O.: *Ein Verfahren der Variationsrechnung das Minimum eines Integral als das Maximum eines anderen Ausdruckes darzustellen*, Göttinger Nachrichten (1929), pp. 13–20.
8. Diaz, J. B. and Weinstein, A.: *Schwarz' inequality and the methods of Rayleigh-Ritz and Trefftz*, J. Math. Phys., **3** (1947), pp. 133–136.
9. Diaz, J. B.: *Upper and lower bounds for quadratic functionals*, Proceedings of Symposium on Spectral Theory and Differential Problems, Oklahoma A. and M. (1950), pp. 279–289.
10. Diaz, J. B.: *Upper and lower bounds for quadratic functionals*, Collectanea Math. **4** (1951), pp. 3–50.
11. Prager, W. and Synge, J. L.: *Approximations in elasticity based on the concept of function space*, Quart. Appl. Math., **5** (1947), pp. 241–269.
12. Payne, L. E. and Weinberger, H. F.: *Upper and lower bounds for harmonic functions, Dirichlet integrals and biharmonic functions*, OSR-TN-54-21, University of Maryland (1954).
13. Trefftz, E.: *Konvergenz und Fehlerschätzung beim Ritzschen Verfahren*, Math. Ann. **100** (1928), pp. 502–521.
14. Weber, C.: *Eingrenzung von Verschiebungen mit Hilfe der Minimalsätze*, Z. Angew. Math. Mech., **22** (1942), pp. 126–130.
15. Diaz, J. B. and Greenberg, H. J.: *Upper and lower bounds for the solution of the first biharmonic boundary value problem*, J. Math. Phys., **27** (1948), pp. 193–201.
16. Greenberg, H. J.: *The determination of upper and lower bounds for the solution of the Dirichlet problem*, J. Math. Phys., **27** (1948), pp. 161–182.
17. Maple, C. G.: *The Dirichlet problem-bounds at a point for the solution and its derivatives*, Quart. Appl. Math., **8** (1950), pp. 213–228.
18. Synge, J. L.: *Upper and lower bounds for the solutions of problems in elasticity*, Proc. Roy. Irish Acad., **53** (1950), pp. 41–64.
19. Synge, J. L.: *Pointwise bounds for the solutions of certain boundary-value problems*, Proc. Roy. Soc. A., **208** (1951), pp. 170–175.
20. Slobodyanskii, M. G.: *Estimates of the error of an approximate solution in linear problems reducing to variational ones, and their application to the determination of two sided approximations in static problems in elasticity*, Akad. Nauk. SSSR Prikl. Math. Meh., **16** (1952), pp. 449–464.
21. Slobodyanskii, M. G.: *Estimate of the error of the quantity sought for in the solution of linear problems by a variational method*, Dok. Akad. Nauk. SSSR (N. S.), **86** (1952), pp. 243–246.
22. Washizu, K.: *Bounds for solutions of boundary value problems in elasticity*, J. Math. Phys., **32** (1953), pp. 117–128.

23. FICHERA, G.: *On some general integration methods employed in connection with linear differential equations*, J. Math. Phys., **29** (1950), pp. 59–68.
24. MIRANDA, C.: *Formule de maggiorazione e teorema di esistenza per le funzioni biarmoniche di due variabili*, Gionale di Mat. di Battaglini, **75** (1948–49).
25. BRELOT, M.: *Sur l'approximation et la convergence dans la theorie des fonctions harmoniques ou holomorphes*, Bull. Soc. Mat. France, **73** (1945), pp. 55–70.
26. DENY, J.: *Sur l'approximation des fonctions harmoniques*, Bull. Soc. Mat. France, **73** (1945), pp. 71–73.
27. MCMAHON, J.: *Lower bounds for electrostatic capacity of a cube*, Proc. Roy. Irish Acad. A, **55** (1953), pp. 133–167.
28. DIAZ, J. B. AND WEINSTEIN, A.: *The torsional rigidity and variational methods*, Amer. J. Math., **70** (1948), pp. 107–116.
29. COURANT, R. AND HILBERT, D.: *Methoden der Mathematischen Physik*, Berlin, Vol. 1 (1937).
30. PICONE, M.: *Sulla convergenza delle successioni di funzioni iperarmoniche*, Bull. Math. Soc. Roum. Sci., **38** (1936).
31. FICHERA, G.: *Teoremi de completezza connessi all'integrazione dell'equazione* $\Delta_4 u = f$. Gior. Mat. Batt., **37** (1947).
32. NICOLESCO, M.: *Les fonctions polyharmoniques*, Hermann, Paris (1936).

INSTITUTE FOR FLUID DYNAMICS AND APPLIED MATHEMATICS
UNIVERSITY OF MARYLAND

(Received March 24, 1954)

# ON THE DYNAMICS OF SMALL VAPOR BUBBLES IN LIQUIDS*

S. A. Zwick and M. S. Plesset

**I. Introduction.** When a vapor bubble in a liquid changes size, evaporation or condensation of the vapor takes place at the surface of the bubble. Because of the latent heat requirement of evaporation, a change in bubble size must therefore be accompanied by a heat transfer across the bubble wall, such as to cool the surrounding liquid when the bubble grows (or heat it when the bubble becomes smaller). Since the vapor pressure at the bubble wall is determined by the temperature there, the result of a cooling of the liquid is a decrease of the vapor pressure, and this causes a decrease in the rate of bubble growth. A similar effect occurs during the collapse of a bubble which tends to slow down the collapse. In order to obtain a satisfactory theory of the behavior of a vapor bubble in a liquid, these heat transfer effects must be taken into account.

In this paper, the equations of motion for a spherical vapor bubble will be derived and applied to the cases of a bubble expanding in superheated liquid and a bubble collapsing in liquid below its boiling point. Because of the inclusion of the heat transfer effects, the equations are nonlinear, integro-differential equations. In the case of the collapsing bubble, large temperature variations occur; therefore, tabulated vapor pressure data were used, and the equations of motion were integrated numerically. Analytic solutions are obtainable for the case of the expanding bubble if the period of growth is subdivided into several regimes and the simplifications possible in each regime are utilized. The growth is considered here only during the time that the bubble is small. An asymptotic solution of the equations of motion, valid when the bubble becomes large (i.e. observable), has been presented previously, together with experimental verification.[1]

We shall be specifically concerned in the following discussion with the dynamics of vapor bubbles in water. This restriction was made for convenience only, since the theory is applicable without modification to many other liquids.

**II. Formulation of the Problem.** The vapor bubble is taken to be spherical. The radial velocity of the bubble wall is considered to be negligible in comparison with sonic velocity in either the liquid or the vapor so that the water will be assumed incompressible, and it will be further assumed that conditions of dynamic and thermal equilibrium exist in the vapor so that the vapor pressure within the bubble will be uniform and equal to the equilibrium vapor pressure of the liquid at temperature of the bubble wall. Viscosity and gravity effects are ignored.† Then in terms of the external pressure $P_0$, the vapor pressure $p_v$ of the fluid at the bubble wall, the density of the water $\rho$ and the surface tension $\sigma$, the radius $R(t)$ of the bubble wall satisfies the differential equation[2]

* This study was supported by the U. S. Office of Naval Research.

[1] M. S. Plesset and S. A. Zwick, J. Appl. Phys. **25**, 493 (1954).

† The justification for these physical assumptions will be found in Ref. 1.

[2] H. Lamb, Hydrodynamics (Dover Pub., New York, 1945). See also M. S. Plesset, J. Appl. Mech. **16**, 277 (1949).

$$R\frac{d^2R}{dt^2}+\frac{3}{2}\left(\frac{dR}{dt}\right)^2=\frac{p_v-P_0}{\rho}-\frac{2\sigma}{\rho R}, \tag{1}$$

with initial conditions

$$R(0)=R_0, \qquad (dR/dt)_{t=0}=0. \tag{2}$$

Unlike the vapor, the liquid cannot be assumed to be in thermal equilibrium. The heat problem in the liquid has been discussed in a previous paper,[3] where a solution was obtained under the assumption that significant heat exchange takes place only in a thin layer of liquid surrounding the bubble wall. If the liquid contains a uniformly distributed heat source per unit volume of magnitude $\eta(t)$, with $\eta(0)=0$, the temperature at the bubble wall is given by*

$$T=T_0+\frac{D}{k}\eta(t)-\left(\frac{D}{\pi}\right)^{\frac{1}{2}}\int_0^t\frac{R^2(x)(\partial T/\partial r)_{r=R(x)}}{\left\{\int_x^t R^4(y)\,dy\right\}^{\frac{1}{2}}}\,dx, \tag{3}$$

corresponding to an initial temperature $T_0$ and a temperature

$$T_\infty=T_0+(D/k)\eta(t) \tag{4}$$

in the liquid far from the bubble wall. Here $D$ is the thermal diffusivity and $k$ the thermal conductivity of the liquid. The temperature gradient in the liquid at the bubble wall, which appears in (3), is obtained from the boundary condition

$$4\pi kR^2(\partial T/\partial r)_{r=R}=(4\pi/3)L\,d(\rho' R^3)/dt, \tag{5}$$

relating the heat flow across the bubble wall to the rate of heat transfer into the bubble due to evaporation or condensation, under the assumption of thermal equilibrium within the bubble. $L$ is the latent heat of evaporation of the water and $\rho'$ the density of vapor within the bubble.

The parameters $\rho$, $\sigma$, $D$, $k$, and $L$ appearing above will be taken as constant, and equal to their values at the initial temperature $T_0$. The vapor density $\rho'$ will be allowed to vary if significant temperature variations occur. The equilibrium vapor pressure $p_v$ is assumed to be a known function of $T$.

In terms of the constants

$$\alpha^2=\frac{2\sigma}{\rho R_0^3}, \qquad c=\frac{L\rho_0' R_0}{3k}\left(\frac{\alpha D}{\pi}\right)^{\frac{1}{2}}, \tag{6}$$

where $\rho_0'=\rho'(T_0)$, we may define a set of dimensionless parameters

$$p=R^3/R_0^3, \qquad u=(\alpha/R_0^4)\int_0^t R^4(y)\,dy,$$
$$\phi=R_0(P_0-p_v)/2\sigma, \qquad r=\rho'(T)/\rho_0', \tag{7}$$

[3] M. S. Plesset and S. A. Zwick, J. Appl. Phys. **23**, 95 (1952).

* See Ref. 3, Eq. (20).

in terms of which the system of equations to be solved becomes

$$\frac{1}{6}\frac{d}{dp}\left[p^{7/3}\left(\frac{dp}{du}\right)^2\right] + \frac{1}{p^{1/3}} + \phi = 0, \tag{8}$$

$$\phi = \phi(T), \tag{9}$$

$$T = T_0 + \frac{D}{k}\eta - c\int_0^u \frac{d(rp)/dv}{(u-v)^{\frac{1}{3}}}\,dv, \tag{10}$$

$$r = r(T). \tag{11}$$

The initial conditions for (8) are

$$p = 1, \qquad dp/du = 0 \qquad \text{for } u = 0. \tag{12}$$

The physical quantities we eventually wish to find are then given by

$$t = \frac{1}{\alpha}\int_0^u \frac{dv}{p^{4/3}}, \qquad R(t) = R_0 p^{1/3}, \qquad \frac{dR}{dt} = \frac{\alpha R_0}{3}\,p^{2/3}\frac{dp}{du},$$
$$T = T_0 + \frac{D}{k}\eta - c\int_0^u \frac{d(rp)/dv}{(u-v)^{\frac{1}{3}}}\,dv. \tag{13}$$

**III. The Expanding Bubble.** The vapor bubble which is to grow in superheated water must at some stage of the superheat be in unstable dynamic equilibrium under the effects of surface tension, vapor pressure and external pressure. Further superheating of the liquid increases the vapor pressure in the bubble and upsets the equilibrium. We shall idealize the actual physical situation by assuming the bubble to be initially in equilibrium in the shape of a sphere; then the condition for equilibrium at the time of release of the bubble $t = 0$ is that $(dR/dt)_0 = (d^2R/dt^2)_0 = 0$, and hence by Eq. (1) that

$$p_v(T_0) - P_0 = 2\sigma/R_0\,. \tag{14}$$

For a given initial temperature $T_0$ and external pressure $P_0$, Eq. (14) fixes the equilibrium radius of the bubble. By Eq. (7), this is equivalent to the condition

$$\phi(T_0) = -1. \tag{15}$$

As the bubble grows, the temperature at the bubble wall decreases toward the boiling point. Inasmuch as water will support only a few degrees of superheat under ordinary conditions, the temperature variation involved in the growth is small and an approximate expression for the dependence of vapor pressure on temperature will suffice. For $P_0 = 1$ atm, a close fit to equilibrium vapor pressure data for water can be obtained between 100°C and 110°C by taking

$$[p_v(T) - P_0]/\rho = A(T - 100), \tag{16}$$

with $A = 40{,}800$ c.g.s. units. By combining (16) with (6), (7), (10), and (14) or (15), we obtain for $\phi$ the relation

$$-\phi(T) = 1 + \frac{AD}{R_0^2\alpha^2 k}\eta - \frac{Ac}{R_0^2\alpha^2}\int_0^u \frac{d(rp)/dv}{(u-v)^{\frac{1}{3}}}\,dv. \tag{17}$$

The choice of $\eta(t)$ depends on the physical situation being described, but in any case the $\eta$ term in (17) is extremely small, and affects the bubble directly only for a minute portion of its growth: the $\eta$ term serves to upset the initial equilibrium. For a temperature rise of 1°C/min in the liquid, this term is of order $10^{-8}$, and it may be neglected once the bubble growth has begun. To fix the model, assume that

$$(D/k)\eta(t) = at, \tag{18}$$

corresponding (cf. Eq. (4)) to a temperature rise of 1°C in $(1/a)$ sec in the liquid far from the bubble. Then from (18), (13),

$$(AD/R_0^2\alpha^2k)\eta(t) = \gamma \int_0^u p^{-4/3}\, dv, \tag{19}$$

where the constant

$$\gamma = aA/R_0^2\alpha^3. \tag{20}$$

In keeping with the above discussion, (19) may be approximated by

$$(AD/R_0^2\alpha^2k)\eta(t) = \gamma u. \tag{21}$$

Because of the small temperature range occurring for the expanding bubble, we shall further approximate $r$ in Eq. (17) by unity, and write Eq. (17) as*

$$-\phi(T) = 1 + \gamma u - \mu \int_0^u \frac{p'(v)\, dv}{(u - v)^{\frac{1}{2}}}, \tag{22}$$

where $p'(v) = dp/dv$ and

$$\mu = Ac/R_0^2\alpha^2. \tag{23}$$

The system of equations thus simplifies to

$$\frac{1}{6}\frac{d}{dp}[p^{7/3}p'^2] = 1 - \frac{1}{p^{1/3}} + \gamma u - \mu \int_0^u \frac{p'\, dv}{(u - v)^{\frac{1}{2}}}$$
$$p' = 0, \qquad p = 1 \quad \text{for} \quad u = 0. \tag{24}$$

A solution to Eq. (24) will be found in three parts, corresponding to three overlapping phases of bubble growth, which may be labeled the "delay period", "early phase", and "intermediate phase".

*Delay Period.* Since the bubble growth starts from equilibrium, put

$$p = 1 + w(u), \tag{25}$$

and assume that initially $w(u)$ and its derivatives are small. On neglecting the

* It is possible, by using the formulas for the temperature given below and a table of equilibrium vapor pressure values, to estimate what the variation in the dimensionless vapor density $r$ would have been, had it been allowed to vary, and to compare this variation with that of $p$. It was found that the ratio $(r'/r)/(p'/p)$ remains less than 5% at any time for the situation considered here.

second or higher powers of $w$, $w'$, $\cdots$, or products of such terms, we may rewrite Eq. (24) in an approximate (linearized) form as

$$w''(u) - w(u) = 3\gamma u - 3\mu \int_0^u \frac{w'(v)\,dv}{(u-v)^{\frac{1}{2}}} \tag{26}$$

$$w(0) = w'(0) = 0.$$

If we now put

$$y(s) = \int_0^\infty e^{-su} w(u)\,du \tag{27}$$

and take the Laplace transform of (26), we obtain for $y(s)$ the equation

$$s^2 y(s) - y(s) = (3\gamma/s^2) - 3\mu s y(s)(\pi/s)^{\frac{1}{2}},$$

whence

$$y(s) = \frac{3\gamma}{s^2} \frac{1}{s^2 - 1 + 3\mu(\pi s)^{\frac{1}{2}}}. \tag{28}$$

In order to match a later solution, we shall be mainly interested in the asymptotic form of the solution of (26) as $u \to \infty$, obtainable from the expansion of (28) about the singularities of $y(s)$. It is possible to obtain a solution to (26) in closed form by the means described below, and also to find a series expansion of $w(u)$ in powers of $u$ from a Laurent expansion of (28) about $s = \infty$, although these will not be used here.

The roots $s^{\frac{1}{2}} = \beta^{\frac{1}{2}}$, say, of

$$s^2 + 3\mu(\pi s)^{\frac{1}{2}} - 1 = 0 \tag{29}$$

correspond to simple poles of $y(s)$. Eq. (28) may therefore be expanded in partial fractions, using the factors indicated in (29); for a given root $s^{\frac{1}{2}} = \beta^{\frac{1}{2}}$, we obtain terms of the form

$$\frac{1}{s^2(s^{\frac{1}{2}} - \beta^{\frac{1}{2}})} = \frac{1}{\beta^2}\left[\frac{1}{(s^{\frac{1}{2}} - \beta^{\frac{1}{2}})} - \frac{\beta^{3/2} + \beta s^{\frac{1}{2}} + s\beta^{\frac{1}{2}} + s^{3/2}}{s^2}\right] \tag{30}$$

multiplied by constant complex coefficients. By the use of the Laplace inversion integral it may be shown that

$$L^{-1}\left[\frac{1}{s^{\frac{1}{2}} - \beta^{\frac{1}{2}}}\right] = \frac{1}{(\pi u)^{\frac{1}{2}}} + \beta^{\frac{1}{2}} e^{\beta u}[1 + \operatorname{erf}(\beta u)^{\frac{1}{2}}] \tag{31}$$

for all complex $\beta^{\frac{1}{2}}$, and hence that for

$$\pi/4 < |\arg \beta^{\frac{1}{2}}| < \pi,$$

(31) vanishes as $u \to \infty$. It follows that the behavior of $w(u)$ as $u \to \infty$ is determined by the singularities of $y(s)$ for which

$$|\arg s^{\frac{1}{2}}| < \pi/4. \tag{32}$$

Actually, there is but one root $s^{\frac{1}{2}} = \beta^{\frac{1}{2}}$ of (29) satisfying condition (32) for $0 < \mu < \infty$, and this root is real.

The residue of $y(s)$ at $s = \beta$ is given by

$$\frac{3\gamma}{\beta^2} \cdot \frac{1}{2\beta + (3\mu\pi^{\frac{1}{2}}/2\beta^{\frac{1}{2}})},$$

or, since $\beta$ satisfies (29), by

$$(3\gamma/\beta^2)2\beta/(3\beta^2 + 1).$$

Hence as $s \to \beta$,

$$y(s) \sim \frac{6\gamma}{\beta(3\beta^2 + 1)} \frac{1}{s - \beta},$$

so that as $u \to \infty$,

$$w(u) \sim \frac{6\gamma}{\beta(3\beta^2 + 1)} e^{\beta u}, \tag{33}$$

where, again, $\beta$ is that root of (29) satisfying (32). Alternatively, Eq. (33) may be written in the form

$$u \sim \frac{1}{\beta} \ln \left[ \frac{\beta(3\beta^2 + 1)}{6\gamma} \cdot w \right] \quad \text{as} \quad w \to \infty. \tag{34}$$

Since the Laplace transform

$$L[T - T_\infty] = -cL\left[ \int_0^u \frac{p'\, dv}{(u - v)^{\frac{1}{2}}} \right] = -c(\pi s)^{\frac{1}{2}} y(s)$$

is asymptotic to $-c(\pi\beta)^{\frac{1}{2}} y(s)$ as $s \to \beta$, we have that

$$T - T_\infty \sim -c(\pi\beta)^{\frac{1}{2}} w(u) \quad \text{as} \quad u \to \infty. \tag{35}$$

Moreover, to the degree of approximation used in the linearization,

$$u = \alpha t, \qquad w = 3\left(\frac{R}{R_0} - 1\right). \tag{36}$$

From (33), (35), (36), we thus obtain the relations

$$R \sim R_0 \left[ 1 + \frac{2\gamma}{\beta(3\beta^2 + 1)} e^{\alpha\beta t} \right], \qquad T \sim T_0 + at - 3c(\pi\beta)^{\frac{1}{2}}[(R/R_0) - 1]. \tag{37}$$

By defining a time $t_0$ by

$$2\gamma/\beta(3\beta^2 + 1) = e^{-\alpha\beta t_0}, \tag{38}$$

we may write the first of Eqs. (37) as

$$R \sim R_0 [1 + e^{\alpha\beta(t - t_0)}], \tag{39}$$

from which we see that the bubble radius remains practically equal to $R_0$ until the time $t \approx t_0 - 1/\alpha\beta$ when it begins to increase, reaching $2R_0$ at time $t \approx t_0$.

TABLE I

*Parameters of the Delay Period*

| $T_0$ | $R_0$ | $t_0$ | $1/\alpha\beta$ | $3c(\pi\beta)^{1/2}$ | $.01\ t_0$ |
|---|---|---|---|---|---|
| 102°C | $1.56 \times 10^{-3}$ cm | $7.34 \times 10^{-4}$ sec | $5.05 \times 10^{-5}$ sec | 1.97°C | $7.34 \times 10^{-6}$ °C |
| 104 | $.75 \times 10^{-3}$ | $8.08 \times 10^{-5}$ | $4.48 \times 10^{-6}$ | 3.30 | $8.08 \times 10^{-7}$ |
| 106 | $.48 \times 10^{-3}$ | $3.09 \times 10^{-5}$ | $1.56 \times 10^{-6}$ | 3.71 | $3.09 \times 10^{-7}$ |

A tabulation of the significant parameters in Eqs. (37), (39) as functions of the initial temperature $T_i$ will be found in Table I, with the choice $a = .01$°C/sec. Because $1/\alpha\beta \ll t_0$, the bubble growth appears to start abruptly near $t = t_0$, rather than at the time of its release $t = 0$. It is for this reason that the interval of time $0 < t \lesssim t_0$ has been called a "delay period".* For given initial conditions $T_0$, $P_0$, the duration of the delay period is completely determined by the choice of the constant $\gamma$, i.e. by the choice of the heat source term $\eta(t)$.

In terms of $t_0$, the last of Eqs. (37) may be written

$$T \sim T_0 + at_0(t/t_0) - 3c(\pi\beta)^{\frac{1}{2}}[(R/R_0) - 1]. \tag{40}$$

It is evident from Table I that the heat source term $at_0(t/t_0)$ in (40) becomes negligible in comparison with the other terms near the end of the delay period. It will be omitted from the subsequent development. From a physical standpoint, this means that the observable bubble behavior is independent of the particular conditions initiating its growth.

The asymptotic formulas for the linearized Eq. (26) presented above are accurate over a range roughly defined by

$$u > 1/\alpha, \qquad w \ll 1;$$

or

$$1/\alpha\beta < t < t_0 - (1/\alpha\beta).$$

Because of the smallness of $\gamma$, $w(u)$ increases over the range by a factor of several powers of $10(\approx 10^5)$.

*The Early Phase.* Neglecting the heat source term and using $w = p - 1$ as independent variable, we may write Eq. (24) as

$$\frac{1}{6}\frac{d}{dw}\left[\frac{(1+w)^{7/3}}{(du/dw)^2}\right]' = 1 - \frac{1}{(1+w)^{1/3}} - \mu\int_0^w \frac{dv}{[u(w) - u(v)]^{\frac{1}{2}}}. \tag{41}$$

For small $w$ this reduces to

$$\frac{1}{6}\frac{d}{dw}\left(\frac{du}{dw}\right)^{-2} \approx \frac{1}{3}w - \mu\int_0^w \frac{dv}{[u(w) - u(v)]^{\frac{1}{2}}},$$

which is satisfied by

$$u = (1/\tau)\ln(Cw) \tag{42}$$

* The existence of such a delay period is important for the macroscopic (asymptotic) bubble growth, cf. Ref. (1).

for arbitrary $C$, provided that

$$\tau^2 = 1 - 3\mu\tau^{\frac{1}{2}} \int_0^1 \frac{dv}{(\ln 1/v)^{\frac{1}{2}}}. \tag{43}$$

Since the integral has the value $\pi^{\frac{1}{2}}$, Eq. (43) is identical with Eq. (29). From the discussion of the delay period it is clear that of the various roots of (43) we must choose that one which satisfies condition (32), $\tau = \beta$. In order to match the previous solution we must further set

$$C = \beta(3\beta^2 + 1)/6\gamma$$

in (42) (cf. Eq. 34)).

It is apparent from (42) that the derivative $du/dw$ of the solution $u(w)$ of (41) has a simple pole at $w = 0$, and we are therefore led to try a solution of the form

$$du/dw = (1/\beta w)\,[1 + a_1 w + a_2 w^2 + \cdots],$$

or

$$u = \frac{1}{\beta} \ln\left[\frac{\beta(3\beta^2 + 1)}{6\gamma}\, w\right] + \frac{a_1}{\beta}\, w + \frac{a_2}{2\beta}\, w^2 + \cdots. \tag{44}$$

By substituting (44) into the integral of (41), we obtain

$$\begin{aligned} \int_0^w \frac{dv}{[u(w) - u(v)]^{\frac{1}{2}}} &= w \int_0^1 \frac{dv}{[u(w) - u(vw)]^{\frac{1}{2}}} \\ &= \beta^{\frac{1}{2}} w \left\{ \int_0^1 \frac{dv}{(\ln 1/v)^{\frac{1}{2}}} - \frac{a_1}{2}\, w \int_0^1 \frac{(1 - v)\, dv}{(\ln 1/v)^{3/2}} \right. \\ &\quad \left. + \left[\frac{3a_1^2}{8} \int_0^1 \frac{(1 - v)^2\, dv}{(\ln 1/v)^{5/2}} - \frac{a_2}{4} \int_0^1 \frac{(1 - v^2)\, dv}{(\ln 1/v)^{3/2}}\right] w^2 + \cdots \right\}. \end{aligned} \tag{45}$$

Now, for $Re(\nu)$, $Re(\lambda) > 0$

$$\int_0^1 z^{\lambda - 1} (\ln 1/z)^{\nu - 1}\, dz = \Gamma(\nu)/\lambda^{\nu}. \tag{46}$$

Consider a typical integral in (45), for example that appearing in the coefficient of $w$ within the braces,

$$I = \int_0^1 \frac{(1 - z)\, dz}{(\ln 1/z)^{3/2}}.$$

If the exponent of the $\ln 1/z$ factor were instead $1 - \nu$, then for $Re(\nu) > 0$ Eq. (46) would give for $I$

$$\int_0^1 \frac{(1 - z)\, dz}{(\ln 1/z)^{\frac{1}{2} - \nu}} = [1 - \tfrac{1}{2}^{\nu}]\Gamma(\nu). \tag{47}$$

But it is readily verified that both sides of (47) are regular functions of the complex variable $\nu$ for $Re(\nu) > -1$, the singularity at $\nu = 0$ being only apparent.

Therefore, by the theory of analytic continuation, the equality (47) remains valid for $Re(\nu) > -1$. In particular, for $\nu = -\frac{1}{2}$ (47) gives

$$I = [1 - 2^{\frac{1}{2}}]\Gamma(-\tfrac{1}{2}).$$

The other integrals appearing in (49) may be similarly evaluated.

From (45), (46), we thus obtain

$$3\mu\int_0^w \frac{dv}{[u(w) - u(v)]^{\frac{1}{3}}} = 3\mu w\beta^{\frac{1}{3}}\Big\{\Gamma(\tfrac{1}{2}) - \tfrac{1}{3}a_1(1 - 2^{\frac{1}{3}})\Gamma(-\tfrac{1}{2})w + [\tfrac{2}{9}a_1^2(1 - 2^{3/2} - 3^{3/2} + 4^{3/2})\Gamma(-\tfrac{3}{2}) - \tfrac{1}{4}a_2(1 - 3^{\frac{1}{3}})\Gamma(-\tfrac{1}{2})]w^2 + \cdots\Big\}. \tag{48}$$

By (41), this must equal

$$3 - 3(1 + w)^{-1/3} - \frac{1}{2}\frac{d}{dw}\left[\frac{(1 + w)^{7/3}}{(du/dw)^2}\right] = (1 - \beta^2)w - [\tfrac{2}{3} + \tfrac{1}{3}\beta^2(7 - 6a_1)]w^2 + [\tfrac{14}{27} - 2\beta^2(\tfrac{14}{9} - \tfrac{14}{9}a_1 + 3a_1^2 - 2a_2)]w^3 + \cdots. \tag{49}$$

The parameters $a_1, a_2, \cdots$ are then found by equating the coefficients of corresponding powers of $w$ in (48), (49).

From (13), the time becomes

$$t = \frac{1}{\alpha}\int_0^w \frac{u'(v)\,dv}{(1 + v)^{4/3}} \sim \frac{1}{\alpha\beta}\left\{\ln\left[\frac{\beta(3\beta^2 + 1)}{6\gamma}w\right] + (a_1 - \tfrac{2}{3})w + \tfrac{1}{2}(a_2 - \tfrac{2}{3}a_1 + \tfrac{14}{9})w^2 + \cdots\right\}, \tag{50}$$

the logarithmic term having been chosen to match (34), (36). The temperature at the bubble wall may be found from (48).

Because the above solution depends upon expansions in powers of $w$ of fractional powers of $1 + w$, it is not expected to be valid for $w$ as large as unity.

*Intermediate Phase.* The early phase solution carries the radius up to about twice its initial value. The asymptotic solution[1] becomes valid at a value of the radius several times larger than this, depending on the superheat, and it becomes necessary to find a solution which will join the other two. This "intermediate phase" solution is conveniently obtained as an expansion about one or two intermediate points on the growth curve and is constructed to fit smoothly onto preceding solutions. The asymptotic solution contains an arbitrary parameter which may be chosen to provide a match between it and the intermediate phase solution.

We have shown that for small $u$, $w'(u) = p'(u)$ grows exponentially. For large $u$, the asymptotic solution gives

$$p(u) \sim (2/\pi\mu)u^{\frac{1}{2}} \quad \text{as} \quad u \to \infty,$$

so that $p'(u)$ dies away as $u^{-\frac{1}{2}}$. Hence the derivative must reach a maximum in

the course of bubble growth, and the function $p(u)$ must pass through a point of inflection. We shall obtain an intermediate solution in the form of an expansion of $p(u)$ about the point $u = u_0$ defined by

$$p''(u_0) = 0. \tag{51}$$

Since the early phase solution is not expected to be accurate at this point, the actual value of $u_0$ or $p(u_0)$ is unknown. In order to determine these quantities, we require that the intermediate solution and its derivative shall coincide with values obtained from the early phase solution at a point $u = u_1$, where that solution is accurate.

The differential equation (24) (with the heat source term omitted) may be written

$$pp'' + \tfrac{7}{6}p'^2 = \frac{3}{p^{4/3}}\left[1 - \frac{1}{p^{1/3}} - \mu\int_0^u \frac{p'(v)\,dv}{(u-v)^{\frac{1}{3}}}\right]. \tag{52}$$

Assume that $p''(u)$ has an expansion about $u_0$

$$p''(u_0 + z) = c_1 z + c_2 z^2 + c_3 z^3 + \cdots,$$

so that

$$\begin{aligned} p'(u_0 + z) &= p_0' + \tfrac{1}{2}c_1 z^2 + \tfrac{1}{3}c_2 z^3 + \cdots, \\ p(u_0 + z) &= p_0 + p_0' z + \tfrac{1}{6}c_1 z^3 + \cdots, \end{aligned} \tag{53}$$

where

$$p_0' = p'(u_0), \qquad p_0 = p(u_0).$$

Now, the integral in (52), evaluated at $u + z$, may be written

$$\begin{aligned} \int_0^{u+z} \frac{p'(v)\,dv}{(u+z-v)^{\frac{1}{3}}} &= \int_0^{u+z} p'(u+z-v)\frac{dv}{v^{\frac{1}{3}}} \\ &= \int_0^u p'(u+z-v)\frac{dv}{v^{\frac{1}{3}}} + \int_u^{u+z} p'(u+z-v)\frac{dv}{v^{\frac{1}{3}}}, \\ &= \sum_{k=0}^{\infty}\frac{z^k}{k!}\int_0^u p^{(k+1)}(u-v)\frac{dv}{v^{\frac{1}{3}}} + \int_0^z \frac{p'(v)\,dv}{(u+z-v)^{\frac{1}{3}}}, \end{aligned} \tag{54}$$

valid for sufficiently small $z$. But for small $z$, $p'(z) \approx 0$, so that when (54) is valid, the last integral in (54) is negligible. Thus for $u = u_0$, (54) gives

$$\int_0^{u_0+z} \frac{p'(v)\,dv}{(u_0+z-v)^{\frac{1}{3}}} \approx I_0 + I_1 z + I_2\frac{z^2}{2!} + \cdots, \tag{55}$$

say, where

$$I_k = \int_0^{u_0} \frac{p^{(k+1)}(v)\,dv}{(u_0-v)^{\frac{1}{3}}}. \tag{56}$$

With the aid of (53), (55), we are now able to write an expansion of (52) about

$u = u_0$. By equating the coefficients of powers of $z$ in this expansion, we obtain a set of relations

$$\frac{7}{6}x^2 = \frac{3}{p^{10/3}}\left[1 - \frac{1}{p^{1/3}} - \mu I_0\right], \quad c_1 = \frac{1}{p^{7/3}}\left[-4x + \frac{5}{p^{1/3}}x - 3\mu(I_1 - \frac{4}{3}xI_0)\right],$$

$$\frac{13}{6}c_1x + c_2 = \frac{1}{p^{7/3}}\left[\frac{14}{3}x^3 - \frac{20}{3p^{1/3}}x^2 - 3\mu\left(\frac{1}{2}I_2 - \frac{4}{3}xI_1 + \frac{14}{9}x^2I_0\right)\right],$$

$$\frac{16}{9}c_2x + c_3 = \frac{3}{p^{7/3}}\left\{-\left(\frac{140}{81}x^3 + \frac{2}{9}\frac{c_1}{p}\right) + \frac{1}{p^{1/3}}\left(\frac{220}{81}x^2 + \frac{5}{18}\frac{c_1}{p}\right)\right. \tag{57}$$

$$\left. -\mu\left[\frac{1}{6}I_3 - \frac{2}{3}xI_2 + \frac{14}{9}x^2I_1 - \left(\frac{140}{81}x^3 + \frac{2}{9}\frac{c_1}{p}\right)I_0\right]\right\}, \cdots,$$

where we have written

$$p \text{ for } p_0, \qquad \text{and} \qquad x \text{ for } p_0'/p_0. \tag{58}$$

The differential equation is satisfied (in principle) by Eqs. (57); there remains the problem of matching with the earlier solution. Therefore, suppose the solution to (52) to be known for $u \leqq u_1$, where $u_1 < u_0$. Put

$$J_k = \int_0^{u_1} \frac{p^{(k+1)}(v)\,dv}{(u_0 - v)^{\frac{1}{2}}}, \qquad L_k = \int_{u_1}^{u_0} \frac{p^{(k+1)}(v)\,dv}{(u_0 - v)^{\frac{1}{2}}}, \tag{59}$$

so that

$$I_k = J_k + L_k \tag{60}$$

in (56), and set

$$\delta = u_0 - u_1, \qquad \xi = u_0 - v. \tag{61}$$

Then in $L_0$, for instance, by (53),

$$p'(v) = p'(u_0 - \xi) = px + \tfrac{1}{2}c_1\xi^2 - \tfrac{1}{3}c_2\xi^3 + \tfrac{1}{4}c_3\xi^4 + \cdots,$$

giving

$$L_0 = \int_{u_1}^{u_0} \frac{p'(v)\,dv}{(u_0 - v)^{\frac{1}{2}}} = 2\delta^{\frac{1}{2}}\left(px + \frac{1}{10}c_1\delta^2 - \frac{1}{21}c_2\delta^3 + \frac{1}{36}c_3\delta^4 + \cdots\right).$$

Thus, $I_0 = J_0(\delta) + L_0(\delta)$ may be found from $p, x, c_1, c_2, c_3, \cdots$ if $\delta$ is given, since then

$$J_0(\delta) = \int_0^{u_1} \frac{p'(v)\,dv}{(u_1 + \delta - v)^{\frac{1}{2}}}$$

is known. In a similar manner, we obtain

$$\begin{aligned}
I_0 &= J_0(\delta) + 2\delta^{\frac{1}{2}}(px + \tfrac{1}{10}c_1\delta^2 - \tfrac{1}{21}c_2\delta^3 + \tfrac{1}{36}c_3\delta^4 + \cdots),\\
I_1 &= J_1(\delta) + 2\delta^{\frac{1}{2}}(-\tfrac{1}{3}c_1\delta + \tfrac{1}{5}c_2\delta^2 - \tfrac{1}{7}c_3\delta^3 + \cdots),\\
I_2 &= J_2(\delta) + 2\delta^{\frac{1}{2}}(c_1 - \tfrac{3}{5}c_2\delta + \tfrac{3}{7}c_3\delta^2 + \cdots),\\
I_3 &= J_3(\delta) + 2\delta^{\frac{1}{2}}(2c_2 - 2c_3\delta + \cdots),\\
&\cdots\cdots.
\end{aligned} \tag{62}$$

It can be shown that the functions $J_k(\delta)$ satisfy an approximate relation

$$J_k(z_1) \approx J_k(z_0) - (z_1 - z_0)\left[\frac{p^{(k+1)}(u_1)}{z_0^k} - J_{k+1}(z_0)\right]. \tag{63}$$

Since in the delay period or early phase $p^{(k+1)} \ll p^{(k)}$, we conclude that $J_k(\delta)$ is a slowly varying function of $\delta$.

Assuming that the contributions to $I_0$, $I_1$, $I_2$, $I_3$ from higher powers of $\delta$ than those written in (62) may be neglected, we terminate the expansions as written and substitute them into (57). Together with the conditions

$$\begin{aligned} p'(u_1) &= px + \tfrac{1}{2}c_1\delta^2 - \tfrac{1}{3}c_2\delta^3 + \tfrac{1}{4}c_3\delta^4, \\ p(u_1) &= p - px\delta - \tfrac{1}{6}c_1\delta^3 + \tfrac{1}{12}c_2\delta^4 - \tfrac{1}{20}c_3\delta^5, \end{aligned} \tag{53'}$$

from (53), the equations (57) constitute a system of six simultaneous (nonlinear) equations for the six unknowns $p$, $x$, $c_1$, $c_2$, $c_3$, $\delta$. Inasmuch as one and only one point of inflection of $p(u)$ is known to exist (from physical considerations), these equations have a unique solution.

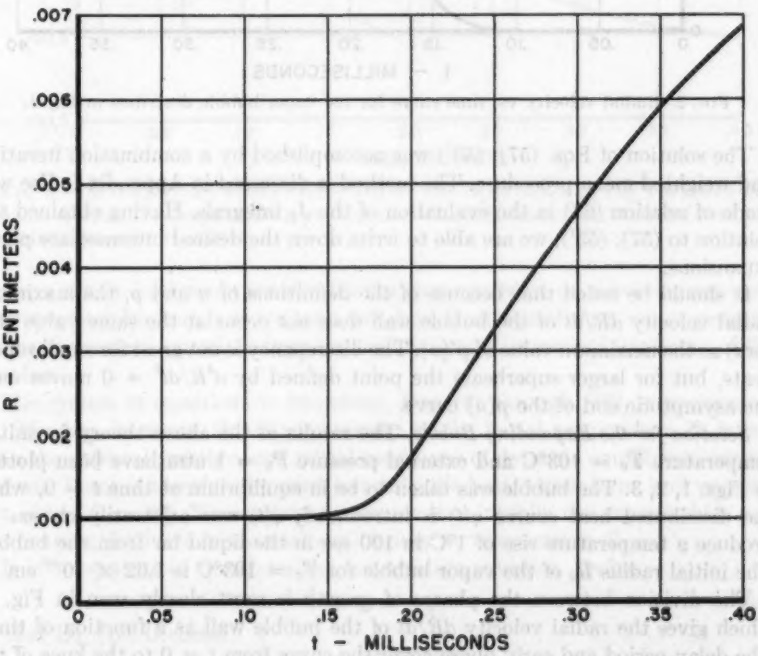



FIG. 1. Theoretical radius vs. time curve for the growth of a spherical vapor bubble in water at 1 atm external pressure, superheated to 103°C. The bubble is originally (at time $t = 0$) in unstable equilibrium. The growth is initiated by the introduction of a uniform heat source in the liquid (e.g. by irradiation), such as to produce a temperature rise in the liquid of 1°C in 100 sec.

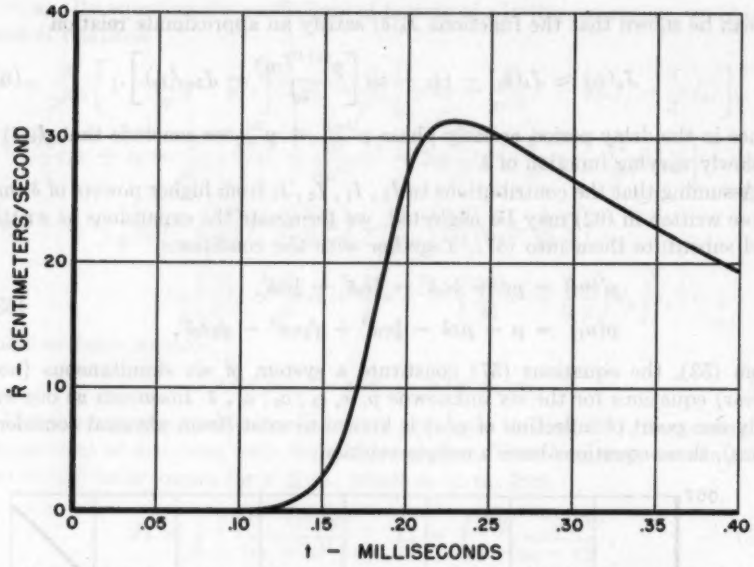



FIG. 2. Radial velocity vs. time curve for the vapor bubble described in Fig. 1.

The solution of Eqs. (57), (53′) was accomplished by a combination iteration and weighted mean procedure. The method is discussed in Appendix I. Use was made of relation (63) in the evaluation of the $J_k$ integrals. Having obtained the solution to (57), (53′), we are able to write down the desired intermediate phase expansions.

It should be noted that because of the definitions of $u$ and $p$, the maximum radial velocity $dR/dt$ of the bubble wall does not occur at the same value of $u$ (or $t$) as the maximum value of $p'(u)$. The discrepancy is not great for small superheats, but for larger superheats the point defined by $d^2R/dt^2 = 0$ moves onto the asymptotic end of the $p(u)$ curve.

*Solution for the Expanding Bubble.* The results of the above theory for initial temperature $T_0 = 103°C$ and external pressure $P_0 = 1$ atm have been plotted in Figs. 1, 2, 3. The bubble was taken to be in equilibrium at time $t = 0$, when the distributed heat source $\eta(t)$ is introduced; $\eta(t)$ was arbitrarily chosen to produce a temperature rise of 1°C in 100 sec in the liquid far from the bubble. The initial radius $R_0$ of the vapor bubble for $T_0 = 103°C$ is $1.02 \times 10^{-3}$ cm.

The division between the phases of growth is most clearly seen in Fig. 2, which gives the radial velocity $dR/dt$ of the bubble wall as a function of time. The delay period and early phase cover the curve from $t = 0$ to the knee of the curve (at about $t = .15$ millisec). The asymptotic solution (treated in Ref. 1) starts just beyond the maximum (at about $t = .25$ millisec).

**IV. The Collapsing Bubble.** Because of the large temperature variations which occur within the bubble which collapses in water below the boiling temper-

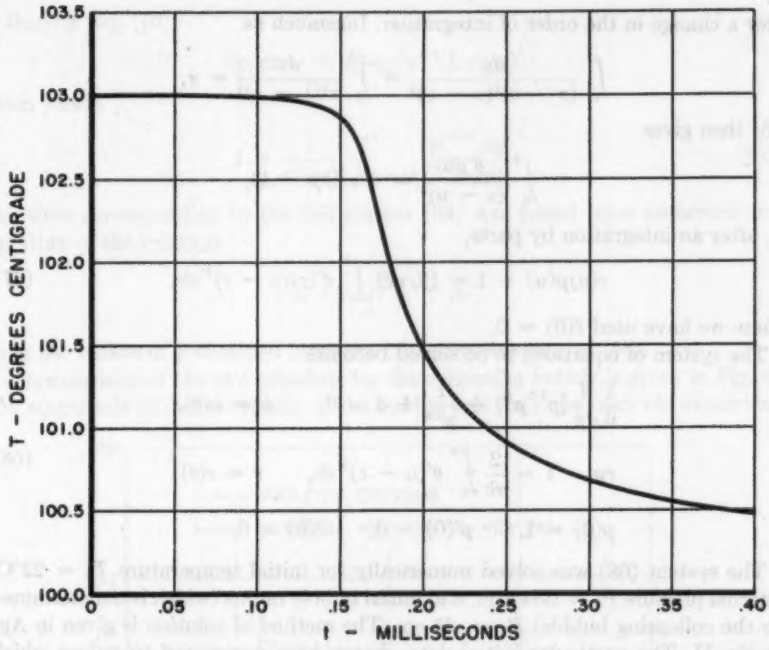



FIG. 3. The temperature at the bubble wall for the vapor bubble of Fig. 1 is shown as a function of time. The temperature in the liquid at a large distance from the bubble remains essentially constant at 103°C.

ature, a simple analytic expression for the vapor pressure or vapor density variations cannot be found. If we take our data from equilibrium vapor pressure tables, we commit ourselves from the beginning to a numerical treatment of the problem.

The system of equations to be solved, Eqs. (8) through (12), is unchanged. In this case, however, the vapor pressure at the initial temperature $T_0$ is less than the external pressure, and initial conditions of dynamic equilibrium cannot prevail. There is therefore no need to retain the heat source term, and we shall put $\eta(t) = 0$ in (10). We continue to assume that initially the vapor bubble and surrounding fluid are in thermal equilibrium at temperature $T_0$.

It is convenient to transform the temperature equation. Define

$$\theta = T - T_0, \tag{64}$$

so that Eq. (10) becomes

$$\theta = -c \int_0^u \frac{(rp)'\, dv}{(u - v)^{\frac{1}{2}}}. \tag{65}$$

If we multiply (65) by $1/(z - u)^{\frac{1}{2}}$ and integrate it from $u = 0$ to $z$, we obtain

$$\int_0^z \frac{\theta(u)\, du}{(z - u)^{\frac{1}{2}}} = -c \int_0^z (rp)'\, dv \int_v^z \frac{du}{(z - u)^{\frac{1}{2}}(u - v)^{\frac{1}{2}}} \tag{66}$$

after a change in the order of integration. Inasmuch as

$$\int_v^u \frac{du}{(z-u)^{\frac{1}{2}}(u-v)^{\frac{1}{2}}} = \int_0^1 \frac{dx}{x^{\frac{1}{2}}(1-x)^{\frac{1}{2}}} = \pi,$$

(66) then gives

$$\int_0^u \frac{\theta\, du}{(z-u)^{\frac{1}{2}}} = -\pi c(rp-1),$$

or, after an integration by parts,

$$r(u)p(u) - 1 = (2/\pi c)\int_0^u \theta'(v)(u-v)^{\frac{1}{2}}\, dv, \tag{67}$$

where we have used $\theta(0) = 0$.

The system of equations to be solved becomes

$$\begin{aligned} &\frac{1}{6}\frac{d}{dp}[p^{7/3}p'^2] + \frac{1}{p^{1/3}} + \phi = 0, \qquad \phi = \phi(\theta), \\ &rp - 1 = \frac{2}{\pi c}\int_0^u \theta'(u-v)^{\frac{1}{2}}\, dv, \qquad r = r(\theta); \\ &p(0) = 1, \qquad p'(0) = 0, \qquad \theta(0) = 0. \end{aligned} \tag{68}$$

The system (68) was solved numerically for initial temperature $T_0 = 22°C$, external pressure $P_0 = .544$ atm, and initial bubble radius (which is undetermined for the collapsing bubble) $R_0 = .25$ cm. The method of solution is given in Appendix II. The particular initial data chosen here correspond to values which have been obtained experimentally.[4]

Although the temperature effects become significant during the collapse, the dynamics of the bubble collapse differ very little from the Rayleigh solution of the problem[5] over most of the collapse. The Rayleigh solution, which neglects heat transfer effects is readily obtained from (68) under the assumption that $\phi$ is constant, and equal to $\phi(T_0)$. We have

$$\frac{1}{6}\frac{d}{dp}[p^{7/3}p'^2] + \frac{1}{p^{1/3}} + \phi_0 = 0, \qquad p(0) = 1, \qquad p'(0) = 0,$$

which yields

$$\tfrac{1}{3}p^{7/3}p'^2 = \phi_0(1-p) + \tfrac{2}{3}(1-p^{2/3}) \tag{69}$$

on integration. Since $\phi_0$ is much larger than $\frac{2}{3}$ ($\phi_0 \approx 10^3$) (69) may be approximated by

$$\tfrac{1}{3}p^{7/3}p'^2 = \phi_0(1-p), \qquad \text{or} \qquad p' = -(6\phi_0)^{\frac{1}{2}}p^{-7/6}(1-p)^{\frac{1}{2}}, \tag{70}$$

where the negative square root of $p'^2$ must be chosen to correspond to the collapsing bubble. From (13)

$$dp/dt = \alpha p^{5/6}p',$$

[4] M. S. Plesset, J. Appl. Mech. **16**, 277 (1949).

[5] Rayleigh, Phil. Mag. **34**, 94 (1917).

so that by Eq. (70),

$$dp/dt = -\alpha(6\phi_0)^{\frac{1}{3}}p^{1/6}(1-p)^{\frac{1}{2}},$$

which yields

$$t = \frac{1}{\alpha(6\phi_0)^{\frac{1}{3}}}\int_{(R/R_0)^3}^{1}\frac{p^{-1/6}\,dp}{(1-p)^{\frac{1}{2}}}. \tag{71}$$

The time corresponding to the full system (68) was found by a numerical integration of the relation

$$t = 1/\alpha\int_0^u p^{-4/3}\,dv,$$

using the values of $p$ obtained from the numerical solution of (68).

A comparison of the two solutions for the collapsing bubble is given in Fig. 4. The magnitude of radial velocity of the bubble wall obtained from the numerical

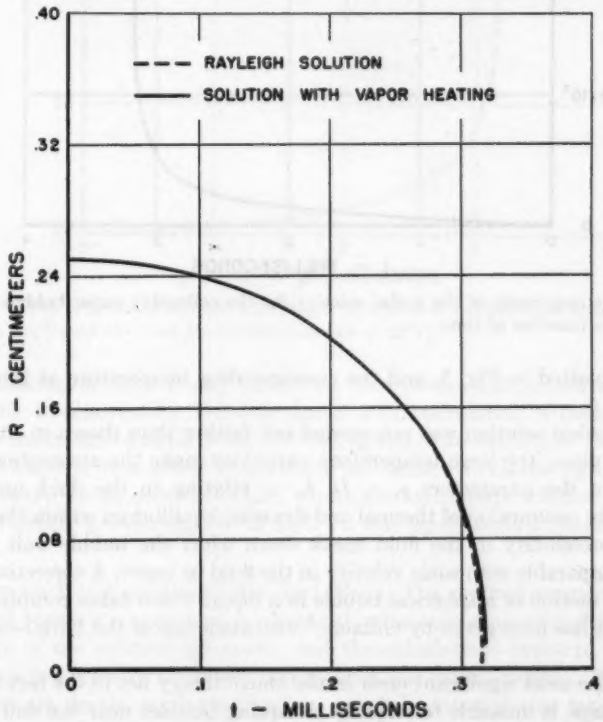



FIG. 4. The radius vs. time curve predicted for the collapsing bubble by the theory developed in the text (solid curve in the figure) is compared with that predicted by the Rayleigh theory (dashed curve). The vapor bubble considered has an initial radius of 0.25 cm; it collapses in water at 22°C with an external pressure of 0.544 atm.

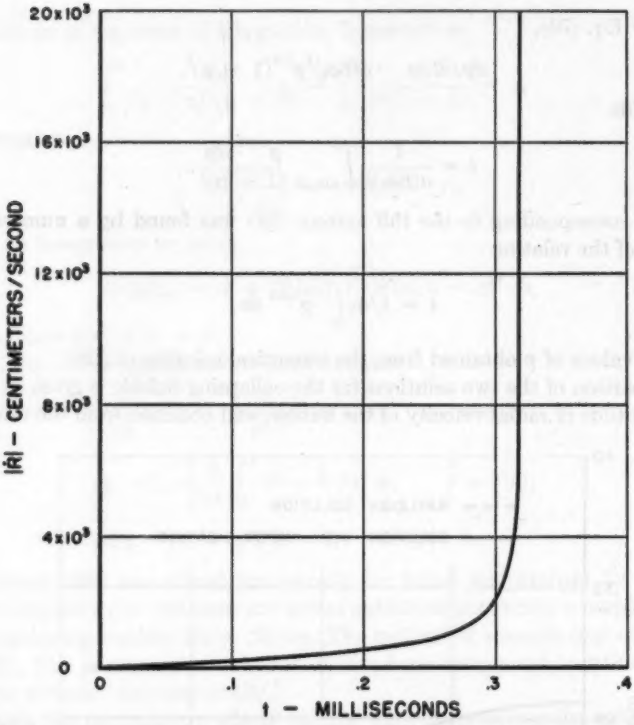



FIG. 5. The magnitude of the radial velocity for the collapsing vapor bubble of Fig. 4 is plotted as a function of time.

solution is plotted in Fig. 5, and the corresponding temperature at the bubble wall in Fig. 6.

The numerical solution was not carried out farther than shown in Figs. 5, 6. In the first place, the large temperature variations make the assumption of the constancy of the parameters $\rho$, $\sigma$, $D$, $k$, $\cdots$ relating to the fluid untenable. Secondly, the assumption of thermal and dynamic equilibrium within the bubble and incompressibility in the fluid break down when the bubble wall velocity becomes comparable with sonic velocity in the fluid or vapor. A correction to the equation of motion of a spherical bubble in a liquid which takes compressibility into account has been given by Gilmore,[6] who made use of the Kirkwood-Bethe hypothesis.

Perhaps the most significant error in the above theory lies in the fact that the spherical shape is unstable for rapidly collapsing bubbles near the end of their collapse.[7] It would therefore be meaningless to carry the numerical solution further, even if the problems mentioned above could be handled.

[6] F. R. Gilmore, Report 26-4, Hydrodynamics Lab., CIT, April 1952.
[7] M. S. Plesset, J. Appl. Phys. **25**, 96 (1954).

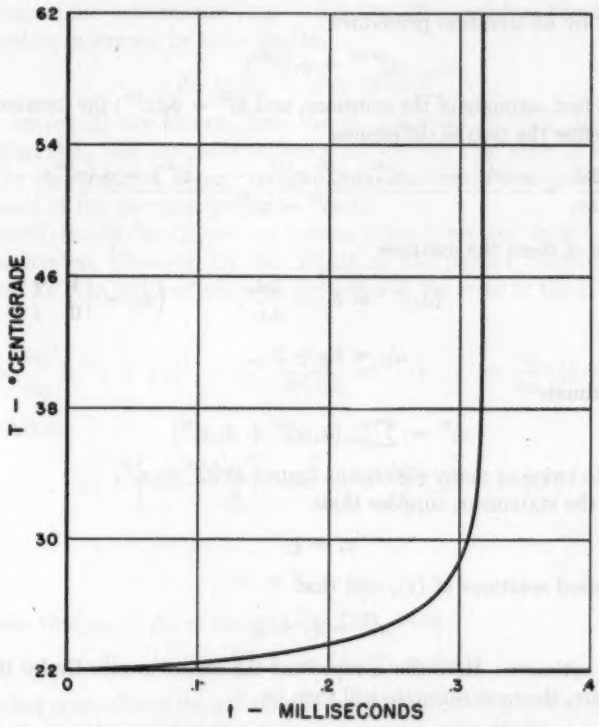



FIG. 6. The temperature at the bubble wall is shown as a function of time. The temperature in the liquid far from the bubble remains at 22°C.

**V. Conclusion.** The theory presented here covers bubble behavior not yet accessible to observation. For this reason, a full tabulation of numerical results for various initial conditions has not been given. The validity of the theory can only be inferred from observations made on bubbles which have grown in superheated water to an observable size, where the asymptotic solution to the equations of bubble growth applies (cf. Ref. 1). The agreement for this range is quite good.

**Appendix I.** *Iteration with Weighted Mean.* The method presented here was designed to solve a system of nonlinear simultaneous equations for which a fair estimate of the solution is known, and the calculations involved are lengthy. It is based essentially on linear interpolation (or extrapolation). When the calculations are simple, methods based on parabolic interpolation may prove more useful.

Consider a set of simultaneous equations of the form (or which can be thrown into the form)

$$x_i = \phi_i(x_1, \cdots, x_n), \qquad i = 1, \cdots, n,$$

to be solved by an iteration procedure, (1)

$$x_i^{(k+1)} = \phi_i(x^{(k)}).$$

Let $x_i^{(1)}$ be a first estimate of the solutions, and $x_i^{(2)} = \phi_i(x^{(1)})$ the iterated second estimate. Define the partial differences

$$\frac{\Delta\phi_i}{\Delta x_j} = \frac{\phi_i(x_1^{(1)}, \cdots, x_{j-1}^{(1)}, x_j^{(2)}, x_{j+1}^{(1)}, \cdots, x_n^{(1)}) - \phi_i(x^{(1)})}{x_j^{(2)} - x_j^{(1)}}, \tag{2}$$

and in terms of them the matrices

$$(\beta_{ij})^{-1} = \delta_{ij} - \frac{\Delta\phi_i}{\Delta x_j}, \qquad \left(\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}\right) \tag{3}$$

$$\alpha_{ij} = \delta_{ij} - \beta_{ij}.$$

Then the estimate

$$x_i^{(3)} = \sum_{j=1}^{n} [\alpha_{ij} x_j^{(1)} + \beta_{ij} x_j^{(2)}] \tag{4}$$

is accurate to twice as many significant figures as $x_i^{(1)}$ or $x_i^{(2)}$.

To prove the statement, suppose that

$$x_i = \xi_i \tag{5}$$

are the required solutions of (1), and that

$$x_i^{(1)} = \xi_i + \epsilon_i \tag{6}$$

are the first estimates. Here the $\epsilon_i$ represent the errors in $x_i^{(1)}$. Under the iteration procedure, the next estimate will then be

$$\begin{aligned} x_i^{(2)} &= \phi_i(x^{(1)}) = \phi_i(\xi + \epsilon) \\ &= \phi_i(\xi) + \textstyle\sum_j \partial\phi_i/\partial x_j(\xi)\epsilon_j + 0(\epsilon^2); \end{aligned}$$

or, since by assumption

$$\phi_i(\xi) = \xi_i, \qquad x_i^{(2)} = \xi_i + \textstyle\sum_j \partial\phi_i/\partial x_j(\xi)\epsilon_j + 0(\epsilon^2). \tag{7}$$

We wish to form a set of weighted means

$$x_i^{(3)'} = \textstyle\sum_j (\alpha_{ij} x_j^{(1)} + \beta_{ij} x_j^{(2)}) \tag{8}$$

which differ from the $\xi_i$ only in $0(\epsilon^2)$. From (6), (7) we find that

$$\begin{aligned} x_i^{(3)} &= \textstyle\sum_j \{\alpha_{ij}[\xi_j + \epsilon_j] + \beta_{ij}[\xi_j + \sum_k \partial\phi_j/\partial x_k(\xi)\epsilon_k]\} + 0(\epsilon^2) \\ &= \textstyle\sum_j (\alpha_{ij} + \beta_{ij})\xi_j + \sum_k (\alpha_{ik} + \sum_j \beta_{ij}\partial\phi_j/\partial x_k)\epsilon_k + 0(\epsilon^2). \end{aligned} \tag{9}$$

The error in $x_i^{(3)}$ will thus be of order $\epsilon^2$ if

$$\alpha_{ij} + \beta_{ij} = \delta_{ij}, \qquad \alpha_{ik} + \textstyle\sum_j \beta_{ij}(\partial\phi_j/\partial x_k)(\xi) = 0. \tag{10}$$

On eliminating $\alpha_{ij}$ from Eqs. (10), we obtain

$$\delta_{ik} = \beta_{ik} - \textstyle\sum_j \beta_{ij}\partial\phi_j/\partial x_k = \sum_j \beta_{ij}(\delta_{jk} - \partial\phi_j/\partial x_k),$$

so that unless the determinant $| \delta_{jk} - (\partial\phi_j/\partial x_k)(\xi) |$ vanishes, the $\beta_{ij}$ are given by the matrix reciprocal to $\delta_{ij} - \partial\phi_i/\partial x_j$ :

$$\beta_{ij} = [\delta_{ij} - (\partial\phi_i/\partial x_j)(\xi)]^{-1}. \tag{11}$$

If the $\partial\phi_i/\partial x_j(\xi)$ are known, then the $\beta_{ij}$ are given by (11), the $\alpha_{ij}$ by the first of Eqs. (10), and the mean estimate $x_i^{(3)}$ by (8). The error in the $x_i^{(3)}$ is of order $\epsilon^2$ by construction. It may be noted that this result does not depend upon convergence of the iteration procedure in (1).

In general, the $\partial\phi_i/\partial x_j(\xi)$ are not known if the $\xi_i$ are not known. These may be approximated, however, by the values of the partial differences $\Delta\phi_i/\Delta x_j$ defined by Eq. (2), without affecting the order of the error in the $x_i^{(3)}$. For, from (2),

$$\frac{\Delta\phi_i}{\Delta x_j} = \frac{\partial\phi_i}{\partial x_j}(x^{(1)}) + \tfrac{1}{2}(x_j^{(2)} - x_j^{(1)})\frac{\partial^2\phi_i}{\partial x_j\,\partial x_j}(x^{(1)}) + \cdots = \frac{\partial\phi_i}{\partial x_j}(\xi) + 0(\epsilon). \tag{12}$$

We now define

$$\beta_{ij} = \left[\delta_{ij} - \frac{\Delta\phi_i}{\Delta x_j}\right]^{-1} = \left[\delta_{ij} - \frac{\partial\phi_i}{\partial x_j}(\xi)\right]^{-1} + 0(\epsilon), \tag{13}$$

and put

$$\alpha_{ij} \equiv \delta_{ij} - \beta_{ij}\,. \tag{14}$$

This means that $\alpha_{ij} + \beta_{ij} = \delta_{ij}$ as in Eq. (10), while

$$\alpha_{ik} + \textstyle\sum_j \beta_{ij}\partial\phi_j/\partial x_k(\xi) = 0(\epsilon).$$

But this last error affects the $x_i^{(3)}$ only in order $\epsilon^2$, according to Eq. (9), and thus does not change the order of the error in the mean estimate.*

The proof breaks down formally when the determinant $\Delta \equiv | \delta_{ij} - \partial\phi_i/\partial x_j(\xi) |$ vanishes. In this case the system (1) may not have a unique solution. For, set

$$\psi_i(x_1\,, \cdots, x_n) \equiv x_i - \phi_i(x_1\,, \cdots, x_n); \tag{15}$$

then $\Delta$ represents the Jacobian of the transformation of variables

$$\psi_i(x_1\,, \cdots, x_n) = \eta_i \tag{16}$$

from $x_i$ to $\eta_i$, evaluated at the point $\eta_i = 0$. The vanishing of the Jacobian signifies, in general, the existence of a family of solutions (depending upon one or more parameters) which passes through $\eta_i = 0$, and correspondingly, unique solutions to (1) cannot exist. If it should happen that $\Delta = 0$ defines an isolated singular point, then it can be shown that the weighted mean procedure based on the $\Delta\phi_i/\Delta x_j$ (for which $\beta_{ij}$ still exists) will converge to that point; however, the convergence in this case goes as $0(\epsilon)$, not $0(\epsilon^2)$, and the method becomes unsuited for numerical computation.

* Clearly, an equally good estimate for the $\partial\phi_i/\partial x_j(\xi)$ would be the $\partial\phi_i/\partial x_j$ evaluated at $x_i^{(1)}$ or $x_i^{(2)}$ if these are readily obtainable, the convergence being relatively insensitive to the choice of $\beta_{ij}$ provided $\alpha_{ij} + \beta_{ij} = \delta_{ij}$.

**Appendix II.** *Numerical Solution for the Collapsing Bubble.* The system (68) may be written

$$pp'' + \frac{7}{6}p'^2 + \frac{3}{p^{4/3}}\left[\frac{1}{p^{1/3}} + \phi\right] = 0, \qquad \phi = \phi(\theta), \tag{1), (2}$$

$$rp - 1 = (2/\pi c)\int_0^u \theta'(v)(u - v)^{\frac{1}{2}}\,dv, \qquad r = r(\theta). \tag{3), (4}$$

In order to obtain a scheme for numerical integration of the system, we first subdivide the range of values of $u$ into intervals defined by the points

$$0 = u_0 < u_1 < \cdots < u_k < \cdots < u_n < \cdots. \tag{5}$$

The intervals corresponding to (5) are in general not equal, but chosen as the integration proceeds.

For convenience, write

$$u_{n+1} - u_n \equiv h, \tag{6}$$

and assume that $p_k = p(u_k)$, $\theta_k = \theta(u_k)$, $r_k = r(\theta_k)$, $\phi_k = \phi(\theta_k)$ are known for $0 \leqq k \leqq n$.

If the interval $u_{k+1} - u_k$ is sufficiently short, $\theta'(u)$ within the interval may be approximated by

$$m_k = \frac{\theta_{k+1} - \theta_k}{u_{k+1} - u_k}. \tag{7}$$

The integral in (3) evaluated at the point $u = u_{n+1}$ may then be estimated as

$$\int_0^{u_{n+1}} \theta'(v)(u_{n+1} - v)^{\frac{1}{2}}\,dv \approx \sum_{k=0}^{n} m_k \int_{u_k}^{u_{k+1}} (u_{n+1} - v)^{\frac{1}{2}}\,dv$$

$$= \tfrac{2}{3}\sum_{k=0}^{n} m_k[(u_{n+1} - u_k)^{3/2} - (u_{n+1} - u_{k+1})^{3/2}].$$

Define

$$I_0 = 0$$

$$I_n = \frac{4}{3\pi c}\sum_{k=0}^{n-1} m_k[(u_{n+1} - u_k)^{3/2} - (u_{n+1} - u_{k+1})^{3/2}]. \tag{8}$$

Then with the use of (6) and (7), Eq. (3) at $u = u_{n+1}$ becomes

$$r_{n+1}p_{n+1} - 1 = I_n + (4h^{\frac{1}{2}}/3\pi c)(\theta_{n+1} - \theta_n). \tag{9}$$

The value of $r_{n+1}$ in (9) may be estimated in terms of $\theta_{n+1}$ by an expansion of $r$ about a value $\theta = \bar{\theta}$ near $\theta_{n+1}$ which uses equilibrium vapor pressure data,* say

$$r_{n+1} = r(\theta_{n+1}) = \bar{r}[1 + \bar{b}_1(\theta_{n+1} - \bar{\theta}) + \bar{b}_2(\theta_{n+1} - \bar{\theta})^2 + \cdots]. \tag{10}$$

* The obvious choice for the first few steps of integration is $\bar{\theta} = 0$, so that $\bar{r} = 1$ initially.

We thus obtain for (9) the equation

$$p_{n+1}\bar{r}[1 + \bar{b}_1(\theta_{n+1} - \bar{\theta}) + \bar{b}_2(\theta_{n+1} - \bar{\theta})^2 + \cdots] = 1 + I_n + (4h^{\frac{1}{2}}/3\pi c)[(\theta_{n+1} - \bar{\theta}) - (\theta_n - \bar{\theta})], \tag{11}$$

in which the only unknowns are $\theta_{n+1}$ and $p_{n+1}$. Eq. (11) is most easily solved for $\theta_{n+1}$ by an iteration procedure based on the equation

$$(\theta_{n+1} - \bar{\theta}) = \frac{p_{n+1}\bar{r} + \dfrac{4h^{\frac{1}{2}}}{3\pi c}(\theta_n - \bar{\theta}) - I_n - 1}{\dfrac{4h^{\frac{1}{2}}}{3\pi c} - p_{n+1}\bar{r}[\bar{b}_1 + \bar{b}_2(\theta_{n+1} - \bar{\theta}) + \cdots]}. \tag{12}$$

To obtain $p_{n+1}$, we use the differential equation (1). At each point $u = u_n$, make the approximations

$$p_n'' = \frac{\left(\dfrac{p_{n+1} - p_n}{u_{n+1} - u_n}\right) - \left(\dfrac{p_n - p_{n-1}}{u_n - u_{n-1}}\right)}{\frac{1}{2}(u_{n+1} - u_{n-1})},$$

$$p_n' = \frac{1}{2}\left[\left(\frac{p_{n+1} - p_n}{u_{n+1} - u_n}\right) + \left(\frac{p_n - p_{n-1}}{u_n - u_{n-1}}\right) - \frac{1}{2}p_n''(u_{n+1} - 2u_n + u_{n-1})\right]. \tag{13}$$

Under these approximations, the formula

$$p(u) = p_n + (u - u_n)p_n' + \tfrac{1}{2}(u - u_n)^2 p_n''$$

is exact for $u = u_{n-1}, u_n, u_{n+1}$. If $u_n - u_{n-1} = u_{n+1} - u_n (\equiv h)$, (13) gives

$$p_n' = \frac{p_{n+1} - p_{n-1}}{2h}, \qquad p_n'' = \frac{p_{n+1} - 2p_n + p_{n-1}}{h^2}, \tag{14}$$

so that the differential equation at $u = u_n$ is approximated by the difference equation

$$p_{n+1}^2 + [\tfrac{24}{7}p_n - 2p_{n-1}]p_{n+1} + [(p_{n-1}^2 - \tfrac{48}{7}p_n^2 + \tfrac{24}{7}p_n p_{n-1}) + 72h^2/7p_n^{1/3}(p_n^{-1/3} + \phi_n)] = 0. \tag{15}$$

Given $\theta_n$, Eq. (15) may be solved for the positive root $p_{n+1}$, $\phi_n = \phi(\theta_n)$ being obtainable from equilibrium vapor pressure data.

In order to keep the differences in $p$ and $\theta$ small and ensure that a positive root $p_{n+1}$ of the difference equation exists, it becomes necessary to decrease $h$ as the numerical integration proceeds. At the step where a new value of $h$ is introduced, the approximations (13) rather than (14) must be used. The difference equation which applies at that step is not (15), therefore, but one obtainable from (13).

To start the integration, a fictitious point $u_{-1} = -u_1$ is used. Corresponding to the initial conditions $p'(0) = 0$ and the approximation (14), we must then choose $p_{-1} = p_1$. Since $p_0 = 1$, the difference equation (15) for $n = 0$ simplifies to the linear equation

$$p_1 = 1 - \tfrac{3}{7}h^2(1 + \phi_0) \tag{16}$$

where $\phi_0 = \phi(\theta_0) = \phi(0)$. The temperature equation (12) becomes

$$\theta_1 = \frac{p_1 - 1}{(4h^{\frac{1}{2}}/3\pi c) - p_1(\bar{b}_1 + \bar{b}_2\theta_1 + \cdots)} \tag{17}$$

for $n = 0$, since $I_0 = \theta_0 = \bar{\theta} = 0$, $\bar{r} = 1$. For sufficiently small $h$, (17) may be approximated by

$$\theta_1 = \frac{p_1 - 1}{(4h^{\frac{1}{2}}/3\pi c) - \bar{b}_1}. \tag{18}$$

It should be noted that $\bar{r}$, $\bar{b}_1$, $\bar{b}_2$, $\cdots$ depend on $\bar{\theta}$, and change with each new expansion of $r(\theta)$. Because these parameters, as well as $h$, may be constant over several integration steps, we have not given them indices which depend on $u$ (i.e. on $n$).

California Institute of Technology
Pasadena, California

(Received March 9, 1954)

# A GENERAL THEORY OF WIDEBAND MATCHING WITH DISSIPATIVE 4 POLES*

By R. LaRosa† and Herbert J. Carlin

**1. Introduction.** If a load impedance is specified it is always possible to introduce a dissipative network in front of this load so that the input is matched to any prescribed tolerance over any prescribed bandwidth. This matching can in fact generally be accomplished with a purely shunt or series connected two terminal network (one-port). However the matching problem takes on a much higher degree of difficulty if the insertion loss of the matching network is taken into account. If the matching network is optimized with respect to insertion loss, then this network must in general be a two-terminal pair structure (two-port). This paper determines the insertion loss limitations for a prescribed bandwidth on such matching two-ports and shows how the results may be used for synthesizing optimum networks.

The general approach to the solution of this problem is to first determine the way in which conditions of physical realizability limit the performance of any matching network without reference to the particular load specified in the problem. Next, the special limitations imposed by the particular load on the performance of the matching network are determined. It is shown that these special limitations are associated with the complex zeros of a transmission function which is defined when the load is represented as a reactance 2-port terminated in a unit resistance.[4]

An important consideration in this study is the choice of an appropriate set of functions which linearly relate excitation and response in the networks under study. Both the tolerance on mismatch and the power transfer to the load are directly expressed as amplitudes of network scattering functions.[5] It can in fact be stated that the scattering coefficient formalism is a key to the solution of the entire problem, and the restrictions on physical realizability expressed in terms of the scattering matrix[6] provide the required tool for obtaining explicit results. It is doubtful whether an impedance or admittance formulation could have led to useful conclusions.

The simplicity of the scattering coefficient formalism makes it possible to predict the power transmission and matching performance of the most efficient, or optimum, matching network without the labor of designing the network in detail. From this it is possible to make general statements concerning the relative merits of lossless and dissipative matching networks.

* This paper contains a portion of the Dissertation for the Degree of Doctor of Electrical Engineering, "Broadband Dissipative Matching Networks" by R. LaRosa, June, 1953. The work reported was sponsored by the Office of Naval Research under Contract No. NOnr-292(00).

† Formerly at Microwave Research Institute of Polytechnic Institute of Brooklyn. Now at Hazeltine Electronics Corporation.

**2. Statement of the Problem.** A generator whose internal impedance is a pure resistance and an arbitrary load are given. The impedance of the load normalized to the resistance of the generator is $Z(p)$ where $p$ is the complex frequency variable $\sigma + j\omega$. The $j\omega$ axis of the $p$ plane corresponds to steady-state sinusoidal waves.

A matching network is used to transmit power from the generator to the load. The function $P(\omega)$ is the power dissipated in the load normalized to the available power of the generator. The normalized system is shown in Fig. 1.

The impedance mismatch at the generator terminals is best specified by the voltage reflection coefficient $S_{11}(p)$ which is related to the normalized input impedance of the matching network by

$$S_{11}(p) = \frac{Z_{in}(p) - 1}{Z_{in}(p) + 1} \tag{1}$$

The magnitude of this function for steady-state sinusoidal excitation is of interest in many engineering problems.

Lossless matching networks are desirable for many applications because they permit all the power, except that which is reflected back toward the generator, to be transmitted to the load. That is, a lossless matching network is characterized by the relation

$$P(\omega) = 1 - | S_{11}(j\omega) |^2 \tag{2}$$

If the reflection coefficient is kept small over the frequency range of interest, $P(\omega)$ will be nearly the total available power of the generator. Over any given frequency range, however, $| S_{11}(j\omega) |$ cannot be made arbitrarily small by means of a lossless matching network. The limitations on the performance of lossless matching networks were first investigated by Bode (Ref. 1) for simple load impedances. Fano (Ref. 2, 3) later extended the analysis to the most general load impedance.

There are at least three reasons why lossy matching networks may be better than lossless matching networks in specific applications. First, a lossless matching network may not be able to provide the desired small $| S_{11}(j\omega) |$ over the

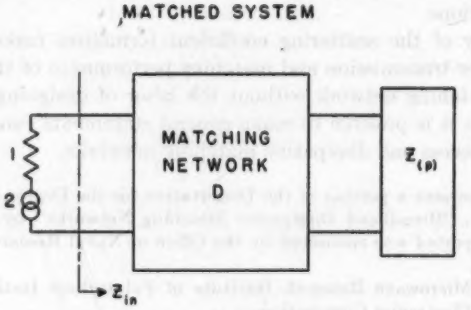



FIG. 1

required frequency range. Second, $P(\omega)$ and $|S_{11}(j\omega)|$ are not independently controlled by means of a lossless matching network. This means, for example, that if $P(\omega)$ is to be decreased in some frequency range for attenuation equalization purposes, the reflection coefficient $S_{11}(j\omega)$ must automatically increase in magnitude in accordance with equation (2). Third, a dissipative matching network might have a simpler form than a lossless network providing the same quality of match. In some cases, the increased simplicity is an advantage which offsets the slight waste of power in the lossy matching network.

In this paper theoretical limitations will be determined which relate $|S_{11}(j\omega)|$ and $P(\omega)$ to the given load impedance when the matching network is composed of lumped inductors, capacitors, resistors, and ideal transformers connected in any fashion without regard to complexity.

**3. Darlington Representation of $Z(p)$.** According to Darlington (Ref. 4) the load impedance $Z(p)$ may be represented by the lossless 2-port (two-terminal-pair network) shown in Fig. 2 terminated in one ohm. This 2-port (labelled $E$

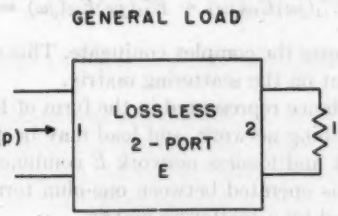



FIG. 2

for identification purposes) is completely specified by a set of three parameters called a scattering matrix.

$$[E(p)] = \begin{bmatrix} E_{11}(p) & E_{12}(p) \\ E_{12}(p) & E_{22}(p) \end{bmatrix} \tag{3}$$

The function $E_{11}(p)$ is the input voltage reflection coefficient when the output port is terminated on one-ohm. This function is determined by the given load impedance according to

$$E_{11}(p) = \frac{Z(p) - 1}{Z(p) + 1} \tag{4}$$

The function $E_{12}(p)$ is the voltage transmission coefficient and is the voltage appearing across a one-ohm termination when a two-volt, one-ohm generator is connected to the opposite end of the network. The magnitude of this transfer function is uniquely determined by the given load impedance because of the relation

$$|E_{11}(j\omega)|^2 + |E_{12}(j\omega)|^2 = 1 \tag{5}$$

which comes from the requirement that the scattering matrix of a lossless network must be unitary for all pure imaginary values of $p$. (Ref. 5)

The function $E_{12}(p)$ is not uniquely determined by a knowledge of the magnitude on the $j\omega$ axis, but it is known from realizability criteria (Ref. 5) that the poles of any scattering function must lie in the interior of the left-hand half of the $p$-plane and that the transfer function of a lossless network may not have zeros in the interior of the left-hand half-plane unless they are paired with corresponding right-hand zeros. (Ref. 7) All poles and zeros must, of course, occur in conjugate pairs or must lie on the real axis.

The $E_{12}(p)$ function for this application should be chosen to satisfy the above requirements with the minimum number of right-hand zeros. If more than the minimum number of zeros is used, a useless complication is introduced into the matching problem.

The function $E_{22}(p)$ is the voltage reflection coefficient at the output port when the input port is terminated in one-ohm. It is uniquely determined from the other two scattering functions according to the relation

$$E_{11}^{*}(j\omega)E_{12}(j\omega) + E_{12}^{*}(j\omega)E_{22}(j\omega) = 0 \tag{6}$$

where the asterisk denotes the complex conjugate. This relation also comes from the unitary requirement on the scattering matrix.

With the load impedance represented in the form of Fig. 2, the complete system of generator, matching network, and load may be represented as in Fig. 3. The matching network and lossless network $E$ combine in tandem to form the single 2-port $S$ which is operated between one-ohm terminations. This over-all network is characterized by a scattering matrix

$$[S(p)] = \begin{bmatrix} S_{11}(p) & S_{12}(p) \\ S_{12}(p) & S_{22}(p) \end{bmatrix} \tag{7}$$

in which the first element is precisely the input reflection coefficient defined in expression (1). Moreover, the power transfer efficiency function $P(\omega)$ is identically the square of the magnitude of the transfer function in the over-all scattering matrix.

Thus, the restrictions imposed by the given load impedance on the functions

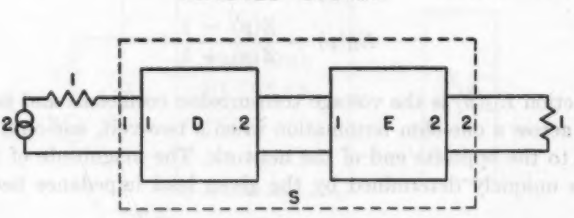



FIG. 3

$|S_{11}(j\omega)|$ and $P(\omega)$ can be found by determining the restrictions on the scattering matrix $[S(p)]$.

**4. Restrictions on** $[S(p)]$. If the matching network $D$ is to be realizable as a combination of inductors, capacitors, resistors, and ideal transformers, the tandem combination of $D$ and $E$ must also satisfy the requirements of physical realizability as outlined by Belevitch (Ref. 6). Thus, elements of $[S(p)]$ must have no poles in the interior of the right-hand half of the $p$-plane. Further, $[1] - [S^*(j\omega)][S(j\omega)]$ must be the matrix of a positive-definite or positive semi-definite hermitian form for $-\infty \leqq \omega \leqq +\infty$.

The latter requirement can be stated in the form of three inequalities:

$$1 - |S_{11}|^2 - |S_{12}|^2 \geqq 0 \qquad (p = j\omega) \qquad (8)$$

$$1 - |S_{22}|^2 - |S_{12}|^2 \geqq 0 \qquad (p = j\omega) \qquad (9)$$

$$(1 - |S_{12}|^2)^2 - |S_{22}|^2 - |S_{11}|^2 + |S_{11}S_{22}|^2 - S_{11}S_{22}S_{12}^{*2} - S_{11}^*S_{22}^*S_{12}^2 \geqq 0 \qquad (p = j\omega) \qquad (10)$$

Satisfaction of inequality (10) and either (8) or (9) insures satisfaction of the remaining inequality.

Certain additional requirements are imposed on the scattering matrix $[S(p)]$ in order that it may represent the tandem combination of lossless 2-port $E$ and a physically realizable matching network $D$. These come from the requirement that the scattering matrix $[D(p)]$ must have no poles in the interior of the right-hand half of the $p$-plane and also from the requirement that $[1] - [D^*(j\omega)][D(j\omega)]$ must be positive-definite or semi-definite for all $\omega$. This latter requirement is already satisfied by satisfaction of inequalities (8), (9), and (10).*

The absence of poles in the function $D_{22}(p)$ is insured by exactly the same integral restrictions on $S_{22}(p)$ as found by Fano (Ref. 2, 3) regardless of whether the matching network is lossy or lossless.

The additional requirement that $S_{11}(p)$ contain all right-hand zeros of $E_{12}(p)$ with at least the same multiplicity is necessary to insure that the functions $D_{11}(p)$ and $D_{12}(p)$ contain no poles in the right-hand half of the $p$-plane.

It has been shown (Ref. 8, 9) that the foregoing restrictions on the scattering matrix $[S(p)]$ are sufficient to insure that $[S(p)]$ can be realized in the form of a realizable passive matching network $D$ in tandem with the given lossless 2-port $E$. Thus we have

THEOREM 1. Given a load impedance $Z(p)$ which is represented in the form of a lossless 2-port $E$ terminated in one-ohm, the necessary and sufficient condi-

* Positive-definiteness or semi-definiteness of the matrix along $j\omega$ insures that the network does not generate sinusoidal waves of any frequency. Conversely, if the network does not generate sinusoidal waves, this matrix is positive-definite or semi-definite. Thus, if the overall combination of $D$ and $E$ does not generate sinusoids, then $D$ alone will not do so because $E$ is lossless and can neither absorb nor generate sinusoidal waves.

tions that $[S(p)]$ represent the scattering matrix of realizable matching network $D$ in tandem with $E$ are:

(a) Matrix $[S(p)]$ should be realizable.

(b) Right-hand (and boundary) zeros of $E_{12}(p)$ should appear in $S_{12}(p)$ with at least the same multiplicity.

(c) At each $n^{th}$ order zero of $E_{12}(p)$ in the interior of the right-hand half of the $p$-plane, $(S_{22}(p) - E_{22}(p))$ must have a zero of order $2n$ at least.

(d) At each boundary zero of $E_{12}(p)$, i.e., where $p_0 = j\omega_0$, $(S_{22}(p) - E_{22}(p))$ has a zero of order $2n$ at least, or order $(2n - 1)$. In the latter case, the first non-zero term in the expansion for $(S_{22}(j\omega) - E_{22}(j\omega))$ about $p = j\omega_0$ is a vector directed toward $E_{22}(j\omega_0)$.*

**5. Restrictions on $| S_{11}(j\omega) |$ and $P(\omega)$.** For practical purposes, Theorem 1 must be interpreted in terms of its effect on functions which describe the performance of the matching network. That is, the restrictions on the magnitude functions $| S_{11}(j\omega) |$ and $P(\omega)$ must be found.

The first step in this process is to translate parts (c) and (d) of Theorem 1 into restrictions on $| S_{22}(j\omega) |$. This has been done by Bode (Ref. 1) and Fano (Ref. 2, 3). These restrictions take the form of weighted integrals taken along the $j\omega$ axis.

$$\int_0^\infty f(\omega) \log \left| \frac{1}{S_{22}(j\omega)} \right| d\omega = F \tag{11}$$

where $f(\omega)$ is the weighting function, and $F$ involves one of the first $(2n - 1)$ coefficients in the expansion for $\log 1/S_{22}(p)$ about a particular right-hand or boundary zero of $E_{12}(p)$. $F$ also involves right-hand zeros of $S_{22}(p)$, some of which are prescribed, and some of which are inserted as adjustable design parameters. The prescribed zeros in $S_{22}$ are the right-hand zeros of $E_{22}$ which are coincident with $E_{12}$ zeros.

In each matching problem there is a set of such integral equations which must be satisfied simultaneously. The number of equations is determined by the number, order, and location of right-hand and boundary zeros of $E_{12}(p)$, i.e. determined entirely by the load. Each equation has its own weighting function and value for $F$. The complete set is satisfied by suitable choice of the $| S_{22}(j\omega) |$ function, by inserting arbitrary right-hand zeros in $S_{22}(p)$, and by changing (in a restricted manner) those $F$ values which correspond to the last controlled coefficient at each boundary zero of $E_{12}(p)$.

If the matching network is lossless, these integrals constitute the solution of the problem because $| S_{22}(j\omega) |$ is simply related to the quantities of interest by

$$| S_{11}(j\omega) | = | S_{22}(j\omega) | \tag{12}$$

$$P(\omega) = | S_{12}(j\omega) |^2 = 1 - | S_{22}(j\omega) |^2 \tag{13}$$

* Parts (c) and (d) are equivalent to corresponding statements in Ref. 2 and Ref. 3 concerning the behavior of $\log 1/S_{11}$ and $\log 1/E_{11}$.

In the general case, where the matching network may contain dissipative elements, the relations between the three magnitude functions are more complicated and are contained in inequalities (9) and (10). Inequality (10) is written here in slightly different form as

$$(1 - |S_{12}|^2)^2 - |S_{22}|^2 - |S_{11}|^2 + |S_{11}S_{22}|^2 - 2\,\text{Real}\,(S_{12}^{*2}S_{11}S_{22}) \geqq 0 \qquad (p = j\omega) \tag{14}$$

The last term can vary between the limits $\pm |S_{12}^2 S_{11}S_{22}|$ depending on the angle of the product. The upper limit is the most favorable for the matching problem because it makes the left-hand side of the inequality as large as possible for a given set of magnitude functions.

To maintain this favorable upper limit it is necessary that

$$S_{12}^{*2}S_{11}S_{22} = -|S_{12}^2 S_{11}S_{22}| \qquad (p = j\omega) \tag{15}$$

This equation may be satisfied without the necessity of inserting right-hand poles in $S_{11}(p)$, $S_{22}(p)$, or $S_{12}(p)$ because $S_{12}(p)$ may be multiplied by a sufficient number of factors of the form

$$\frac{p - p_z}{p + p_z^*} \tag{16}$$

where $p_z$ is located in the right-hand half plane. The only difficulty in satisfying equation (15) arises in making $|S_{11}(j\omega)\cdot S_{22}(j\omega)|$ rational. There are special cases in which this problem is not encountered, i.e. when the matching network is lossless and when $S_{11}$ is identically zero.

In the general case where the simplest functions giving the desired $|S_{11}(j\omega)|$ and $|S_{22}(j\omega)|$ do not have a rational product, it is necessary to approximate one or both of these functions with (usually) more complicated algebraic functions whose product is rational. This makes it more difficult but not impossible to fit rational fractions to the desired magnitude functions when the most favorable extreme of inequality (14) is to be realized. As long as the approximation procedure can be carried out with an arbitrarily small error, this inequality may be written in its most favorable form which, fortunately, contains only magnitudes.

$$(1 - |S_{12}|^2)^2 - |S_{22}|^2 - |S_{11}|^2 + |S_{11}S_{22}|^2 + 2|S_{12}^2 S_{11}S_{22}| \geqq 0 \qquad (p = j\omega) \tag{17}$$

This expression can be factored into

$$\{(1 + |S_{11}|)(1 - |S_{22}|) - |S_{12}|^2\}\{(1 - |S_{11}|)(1 + |S_{22}|) - |S_{12}|^2\} \geqq 0,\ p = j\omega \tag{18}$$

Both factors must either be positive or negative. They cannot both be negative because this would require violating inequality (8) or (9). In order to insure that both factors are positive, it is only necessary to insure that the least positive factor is greater than zero. Thus

If $|S_{11}(j\omega)| > |S_{22}(j\omega)|$

$$(1 - |S_{11}(j\omega)|)(1 + |S_{22}(j\omega)|) - |S_{12}(j\omega)|^2 > 0 \tag{19}$$

or if $|S_{11}(j\omega)| \leqq |S_{22}(j\omega)|$

$$(1 + |S_{11}(j\omega)|)(1 - |S_{22}(j\omega)|) - |S_{12}(j\omega)|^2 \geqq 0 \tag{20}$$

Relation (19) or (20) not only replaces inequality (10), but also insures satisfaction of (8) and (9), so that only a single inequality need be considered in any frequency range. This, of course, is contingent upon satisfaction of equation (15). Thus, we have

THEOREM 2. The input reflection coefficient magnitude $|S_{11}(j\omega)|$ and the power transfer function $P(\omega) = |S_{12}(j\omega)|^2$ can always be obtained with a structure consisting of a physically realizable matching network in tandem with a prescribed lossless 2-port $E$ if:

(a) (1) $P(\omega) \leqq (1 + |S_{11}(j\omega)|)(1 - |S_{22}(j\omega)|)$ if $|S_{11}(j\omega)| \leqq |S_{22}(j\omega)|$

(2) $P(\omega) < (1 - |S_{11}(j\omega)|)(1 + |S_{22}(j\omega)|)$ if $|S_{11}(j\omega)| > |S_{22}(j\omega)|$

(b) $|S_{22}(j\omega)|$ satisfies all the integral restrictions (of Ref. 2, 3) imposed by the lossless 2-port $E$ with all necessary right-hand zeros inserted in the $S_{22}(p)$ function

(c) $P(\omega)$ and the product $|S_{11}(j\omega)S_{22}(j\omega)|$ are rational.

Thus, the limitations on the performance of any matching network are given in a simple form which can be easily interpreted in terms of quantities of direct engineering importance.

**6. Optimum Matching Network.** In a particular matching problem there is some idea or specification of the maximum allowable $|S_{11}(j\omega)|$. Moreover, there is usually a particular shape of the $P(\omega)$ curve appropriate to the given problem, and it is desirable to know what is the maximum allowable scale factor on the $P(\omega)$ curve. The matching network giving this $P(\omega)$ curve with maximum scale factor is here called the optimum matching network.

In order to predict the form of the optimum matching network it is necessary to repeat here the order in which restrictions are imposed on $|S_{22}(j\omega)|$, $|S_{11}(j\omega)|$, and $P(\omega)$.

First, an upper limit is specified for the $|S_{11}(j\omega)|$ function. Second, the shape of the power curve is specified by defining a function $T(\omega)$ such that

$$P(\omega) = kT(\omega) \tag{21}$$

where $k$ is a positive real constant. $T(\omega)$ must be positive, real, rational, and even in $\omega$.

Now, looking at inequality (20), which may be written as

$$P(\omega) \leqq (1 + |S_{11}(j\omega)|)(1 - |S_{22}(j\omega)|) \tag{22}$$

it is evident that when $|S_{22}(j\omega)| \geqq |S_{11}(j\omega)|$ the smallest value for $|S_{22}(j\omega)|$ will give the largest allowable value for $P(\omega)$. This means that $|S_{22}(j\omega)|$ should be as small as possible consistent with the integral restrictions.

There is no advantage, however, in making $|S_{22}(j\omega)|$ smaller than $|S_{11}(j\omega)|$ because then inequality (19) applies instead of (20), and the allowable $P(\omega)$ is reduced as $|S_{22}(j\omega)|$ is decreased. For a fixed $|S_{11}(j\omega)|$ the maximum allowable $P(\omega)$ is obtained when $|S_{22}(j\omega)| = |S_{11}(j\omega)|$. Thus, inequality (20) or (22) is the only one which should be considered when designing a matching network because $|S_{22}(j\omega)|$ should never be smaller than $|S_{11}(j\omega)|$.

With this in mind, inequality (22) shows that if the largest value of $k$ is to be obtained (equation 21) then the upper limit allowed by (22) should be used throughout the entire frequency range. That is, the optimum matching network is characterized by the equation

$$P(\omega) = (1 + |S_{11}(j\omega)|)(1 - |S_{22}(j\omega)|) \tag{23}$$

where

$$|S_{22}(j\omega)| \geqq |S_{11}(j\omega)|$$

Thus, if $|S_{11}(j\omega)|$ and $T(\omega)$ are specified, then the output reflection coefficient function $|S_{22}(j\omega)|$ is completely determined as a function of $k$ by equation (23). The constant $k$ becomes a parameter in the system of integral restrictions on $|S_{22}(j\omega)|$ referred to in connection with Theorem 2. The other design parameters in the system are adjusted to satisfy all equations simultaneously with the largest value of $k$.

In order to satisfy both equations (23) and (15) it is necessary that $|S_{11}(j\omega)|$ be rational or that $|S_{11}(j\omega)| \equiv |S_{22}(j\omega)|$. The latter condition can exist only if the matching network is lossless.

Satisfaction of equation (23) requires that the matrix $[1] - [S^*(j\omega)][S(j\omega)]$ be singular for all $\omega$. This means that the matching network need not contain more than one resistor because its resistance or conductance matrix is singular. This fact does not preclude the use of more than one resistor in the synthesis of a matching network, and it is often simpler to synthesize the network with more than one resistor. The following theorem summarizes the conclusions of this section.

THEOREM 3. The optimum matching network constructed from inductors, capacitors, resistors, and ideal transformers has the following characteristics.

(a) $|S_{11}(j\omega)| \leqq |S_{22}(j\omega)|$

(b) $P(\omega) = kT(\omega) = (1 + |S_{11}(j\omega)|)(1 - |S_{22}(j\omega)|)$

(c) Assuming $T(\omega)$ and $|S_{11}(j\omega)|$ are specified, $k$ is the largest value permitted by the integral restrictions of Reference 2, 3.

(d) $|S_{11}(j\omega)|$ is rational or the matching network is lossless.

(e) It is possible to construct the network with no more than a single resistor.

EXAMPLE 1. A simple load impedance consists of a 1 ohm resistor and inductor $L$ in series as shown in Fig. 4. Assume that a matching network is to be designed

to make $| S_{11}(j\omega) |$ equal to a constant $| S_{11} |$ for all $\omega$ and that the power transfer efficiency is to have the following form.

$$\begin{aligned} T(\omega) &= 1 \qquad 0 \leqq \omega \leqq 1 \\ T(\omega) &= 0 \qquad \omega > 1 \end{aligned} \tag{24}$$

where $T(\omega)$ is defined by equation (21). It is now desired to compute the maximum value of $k$ which can be obtained.

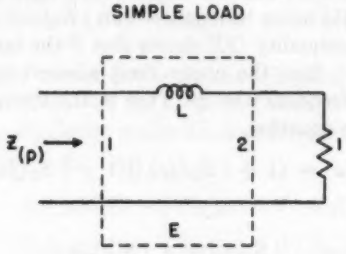



FIG. 4

It will be noted that the simple load of Fig. 4 is already represented in the Darlington form. The lossless network $E$ has a simple zero of $E_{11}(p)$ at infinity. Referring to the Fano work (Reference 2, 3) the single applicable integral restriction on $| S_{22}(j\omega) |$ is

$$\int_0^\infty \log \left| \frac{1}{S_{22}(j\omega)} \right| d\omega = \pi \left( \frac{1}{L} - \sum_s p_s \right) \tag{25}$$

where the $p_s$ are arbitrarily inserted right-hand zeros of $S_{22}(p)$. These zeros are detrimental because they have the same effect as increasing $L$, and since they are not required here, they are left out.

Part (b) of Theorem 3 requires that

$$| S_{22}(j\omega) | = 1 \qquad \omega > 1 \tag{26}$$

and that in the frequency range $0 \leqq \omega \leqq 1$, $| S_{22}(j\omega) |$ should be a constant whose value can be determined from equation (25) as

$$| S_{22}(j\omega) | = \epsilon^{-\pi/L} \qquad 0 \leqq \omega \leqq 1 \tag{27}$$

which means that the constant $k$ is $(1 + | S_{11} |)(1 - \epsilon^{-\pi/L})$ provided $\epsilon^{-\pi/L} \geqq | S_{11} |$. Thus it can be seen that increasing the input mismatch will allow increased power-transfer efficiency to be obtained, with a maximum value of

$$k = (1 - \epsilon^{-2\pi/L}).$$

This maximum value is obtained with a lossless matching network.

EXAMPLE 2. The functions defined for $T(\omega)$ and $| S_{22}(j\omega) |$ in example 1 have sharp discontinuities and a matching network would have to have many elements in order to approximate closely these functions.

To illustrate the procedure of designing an optimum matching network, a simple function will be used to approximate the $T(\omega)$ function defined by (24).

$$T(\omega) = 1/(1 + \omega^2) \tag{28}$$

The input mismatch is again specified to be the constant $| S_{11} |$. Theorem 3 requires that

$$P(\omega) = k/(1 + \omega^2) = (1 + | S_{11} |)(1 - | S_{22}(j\omega) |) \tag{29}$$

The value of $k$ can be determined by solving equation (29) for $| S_{22}(j\omega) |$ and inserting this value in the integral equation (25). It is easier to obtain the value of $k$ by proceeding directly with the synthesis of the matching network using the variable $k$ until examination of the network discloses what the value should be.

The function $S_{22}(p)$ is easily determined from the knowledge that $S_{22}(p)$ should have no right-hand zeros as explained in connection with integral restriction (25), and for realizability no right-hand poles. From equation (29)

$$| S_{22}(j\omega) | = \frac{1 - [k/(1 + S_{11})] + \omega^2}{1 + \omega^2} \tag{30}$$

and the function $S_{22}(p)$ is

$$S_{22}(p) = \frac{\{p + \sqrt{1 - [k/(1 + | S_{11} |)]}\}^2}{(p + 1)^2} \tag{31}$$

The simplest possible $S_{12}(p)$ is

$$S_{12}(p) = k^{\frac{1}{2}}/(p + 1) \tag{32}$$

The function $S_{11}(p)$ is given by equation (15) which must be satisfied by an optimum matching network

$$S_{11}(p) = | S_{11} | \frac{p - \sqrt{1 - [k/(1 + | S_{11} |)]}}{p + \sqrt{1 - [k/(1 + | S_{11} |)]}} \tag{33}$$

The functions defined by (31), (32), and (33) are the elements of the scattering matrix $[S(p)]$ of the matching network with the inductance $L$ in series with its output terminals. The network may be easily synthesized by computing the open-circuit impedance matrix corresponding to $[S(p)]$ and examining its poles. When this is done, it is found that the open-circuit output impedance contains a simple pole at $p = \infty$ which is not contained in the open-circuit input impedance or transfer impedance. This pole corresponds to an inductance whose value is

$$\frac{1}{1 - \sqrt{1 - [k/(1 + | S_{11} |)]}} \tag{34}$$

But the matching network itself must not contain an inductance in series with its output terminals because this would merely add to the inductance

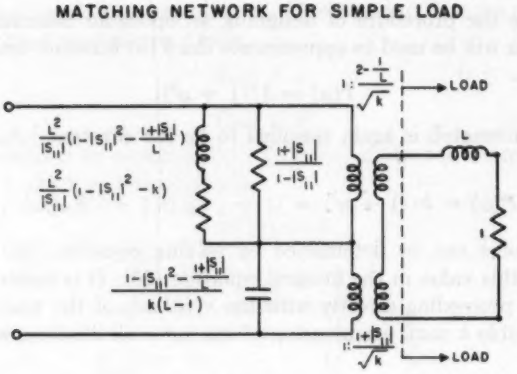



FIG. 5

that is being matched out. Therefore the value given by (34) is exactly the inductance $L$ of the load. Thus the equation

$$L = \frac{1}{1 - \sqrt{1 - [k/(1 + |S_{11}|)]}} \tag{35}$$

relates the power-transfer efficiency and input mismatch to the given load inductance.

It will be noted that equation (35) requires that $L$ be not less than unity. This restriction was not contained in the original specification of the problem but was inserted when the approximating function (28) was chosen for $T(\omega)$.

The complete matching network is shown in Fig. 5. The element values are given in terms of all three parameters even though they are related by equation (35). This results in the simplest expressions for the element values.

In accordance with Part (a) of Theorem 3, the minimum value of $|S_{22}(j\omega)|$ must not be less than $|S_{11}|$, leading to the requirement

$$1 - [k/(1 + |S_{11}|)] \geq |S_{11}| \tag{36}$$

which may be rewritten using (35) to give

$$1 - L^{-1} \geq \sqrt{|S_{11}|} \tag{37}$$

which puts an upper bound on $|S_{11}|$. This situation is similar to that described in connection with example 1 where $|S_{11}|$ was bounded by $\epsilon^{-\pi/L}$.

EXAMPLE 3. Consider the load impedance

$$Z(p) = \frac{2p + 1}{p + 1} \tag{38}$$

whose simplest circuit representation is shown in Fig. 6 and whose Darlington representation is shown in Fig. 7.

The following performance requirements are arbitrarily imposed on the optimum matching network:

$$S_{11}(p) \equiv 0 \tag{39}$$

$$T(\omega) = 1 \qquad 0 \leqq \omega \leqq \infty \tag{40}$$

The value of $k$ achieved by an optimum matching network is to be found. The transfer function

$$E_{12}(p) = -2\,\frac{\sqrt{2}p - 1}{3p + 2} \tag{41}$$

and the output reflection coefficient

$$E_{22}(p) = -\,\frac{p(\sqrt{2}p - 1)}{(3p + 2)(\sqrt{2}p + 1)} \tag{42}$$

both have a simple zero at $p = 1/\sqrt{2}$.

In this case, referring to References 2 and 3, a single integral equation must be satisfied:

$$\int_0^\infty \frac{\log \mid 1/S_{22}(j\omega) \mid}{1 + 2\omega^2}\,\sqrt{2}\,d\omega = \frac{\pi}{2}\log\,(3 + 2\sqrt{2}) \tag{43}$$

Part (b) of Theorem 3 reduces to a simple expression because $S_{11}(p) \equiv 0$

$$k = 1 - \mid S_{22}(j\omega) \mid \tag{44}$$

and when this constant value for $\mid S_{22}(j\omega) \mid$ is substituted in equation (43) the value for $k$ is found to be

$$k = \frac{2 + 2\sqrt{2}}{3 + 2\sqrt{2}} \tag{45}$$

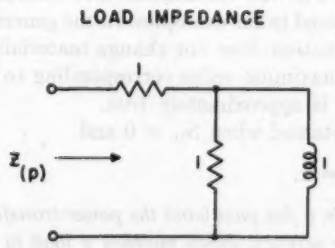



FIG. 6

DARLINGTON REPRESENTATION OF THE LOAD IMPEDANCE OF FIG. 6

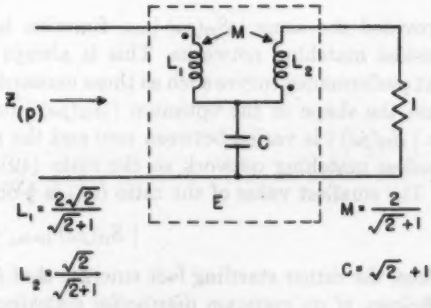



FIG. 7

In practical problems this computation is valuable because it provides a standard with which to compare the power transfer efficiency of simple, practical matching networks.

**7. Special Case of Reflectionless Matching.** Example 3 was an illustration of the special case of reflectionless matching in which the constant-resistance generator operates into a perfect match at all frequencies. The performance of a reflectionless optimum matching network (Ref. 10) is characterized by the equation

$$P(\omega) = kT(\omega) = 1 - |\, S_{22}(j\omega)\,| \tag{46}$$

which is obtained from part (b) of Theorem 3 by introducing the matching constraint $S_{11}(p) \equiv 0$, corresponding to a unit normalized constant input resistance.

It is interesting to compare the performance of optimum dissipative matching networks with optimum lossless networks (the latter designed on the basis of minimum input reflection amplitude over a prescribed band). This is easily done since the limiting factor in both lossy and lossless networks is the amplitude function $|\, S_{22}(j\omega)\,|$ as determined by the integral constraints imposed by the prescribed load. The minimum value of input reflection factor amplitude over a given pass band when lossless matching networks are used is $|\, S_{22}(j\omega)\,|_{\text{min.}}$, and the maximum power-transfer efficiency is:

$$P(\text{Max. Lossless}) = 1 - |\, S_{22}(j\omega)\,|^2_{\text{min.}} \tag{47}$$

Part (b) of Theorem 3 gives the maximum power-transfer efficiency of a dissipative matching network in terms of $|S_{22}(j\omega)\,|_{\text{min.}}$ when $|\, S_{11}(j\omega)\,|$ is prescribed:

$$P(\text{Max. Lossy}) = (1 + |\, S_{11}(j\omega)\,|)(1 - |\, S_{22}(j\omega)\,|_{\text{min.}}) \tag{48}$$

The ratio of power transfer efficiencies when a given load is matched by optimum lossy and lossless networks is:

$$\frac{P(\text{Max. Lossy})}{P(\text{Max. Lossless})} = \frac{1 + |\, S_{11}(j\omega)\,|}{1 + |\, S_{22}(j\omega)\,|_{\text{min.}}} \tag{49}$$

provided the same $|\, S_{22}(j\omega)\,|_{\text{min.}}$ function is optimum for both the lossy and lossless matching networks. This is always true for rectangular and entirely flat performance curves such as those encountered in the examples. In the general case the shape of the optimum $|\, S_{22}(j\omega)\,|$ function does not change materially as $|\, S_{11}(j\omega)\,|$ is varied between zero and the maximum value corresponding to a lossless matching network so the ratio (49) is approximately true.

The smallest value of the ratio (49) is $\frac{1}{2}$ obtained when $S_{11} = 0$ and

$$|\, S_{22}(j\omega)\,|_{\text{min.}} = 1.$$

Thus, the rather startling fact emerges that *in a flat pass band the power-transfer efficiency of an optimum dissipative matching network which matches a load to a real generator impedance without reflections over the infinite frequency band can*

*never be inferior to that of an optimum lossless network by more than* 3 *db.* It should be pointed out that in practical problems $| S_{22}(j\omega) |_{\min.} < 1$, so that the difference in power-transfer efficiency between lossy and lossless matching is invariably less than 3 db.

## REFERENCES


1. H. W. Bode, "Network Analysis and Feedback Amplifier Design" (Book), D. Van Nostrand Company, New York, 1945.
2. R. M. Fano, "Theoretical Limitations on the Broadband Matching of Arbitrary Impedances", Doctor of Science Thesis, Massachusetts Institute of Technology 1947.
3. R. M. Fano, "Theoretical Limitations on the Broadband Matching of Arbitrary Impedances", Jour. Franklin Inst., v 249, n 1, p 57, (Jan. 1950); v 249, n 2 p 139, (Feb. 1950).
4. S. Darlington, "Synthesis of Reactance 4-Poles", Jour. Math. and Physics, Sept. 1939; also Bell Telephone System Monograph B-1186.
5. C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits" (Book), McGraw-Hill Book Company, New York, 1948 (Volume 8 of Radiation Laboratory Series).
6. V. Belevitch, "Synthesis of 2n Terminal Networks", Wireless Engineer, v 28 (1951) p. 128.
7. H. J. Carlin and R. LaRosa, "On the Synthesis of Reactance 4-Poles", Journal of Applied Physics, Vol. 24, n 10, 1336-1337, Oct. 1953.
8. R. LaRosa and H. J. Carlin, "A General Theory of Wideband Matching with Dissipative 4-Poles", Microwave Research Institute of the Polytechnic Institute of Brooklyn, Report R-308-53, PIB-247, Contract No. NOnr-292(00), Feb. 17, 1953.
9. R. LaRosa, "Broadband Dissipative Matching Networks", Doctor of Electrical Engineering Dissertation, Polytechnic Institute of Brooklyn, June, 1953.
10. H. J. Carlin and R. LaRosa, "Broad-Band Reflectionless Matching with Minimum Insertion Loss", Proceedings of the Symposium on Modern Network Synthesis sponsored by Polytechnic Institute of Brooklyn and Office of Naval Research, New York, N.Y., April 1952, p 161.


Polytechnic Institute of Brooklyn

(Received April 27, 1953)

# THE SOLUTION BY ITERATION OF NONLINEAR INTEGRAL EQUATIONS

By Mark Lotkin

**1. Introduction.** It is a well known fact that many physical problems lead to nonlinear integral equations. The following examples serve to illustrate this point.

i. Nyström has shown [1] that the general first boundary value problem in the theory of ordinary differential equations may be reduced to an ordinary differential equation of the type

$$y'' = F(x, y) \qquad -\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2}$$

with the boundary conditions $y(-\frac{1}{2}) = y(\frac{1}{2}) = 0$. This equation, in turn, is equivalent to the integral equation

$$y(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} K(x, s)F(s, y(s))\, ds,$$

$K(x, s)$ representing the Green's function of the boundary value problem.

ii. Boundary value problems of the form

$$\Delta u \equiv u_{xx} + u_{yy} = \exp u$$

with $u = f(x, \xi)$ on the boundary $C$ of the given region $B$, are reducible to integral equations, as follows [2]: Let $w(x, \xi)$ be the solution of the boundary value problem

$$\Delta w = 0, \qquad w = f \text{ on } C.$$

Then the function $y = u - w$ clearly satisfies

$$\Delta y - y \exp w = [\exp (y) - y] \exp w$$

with $y = 0$ on $C$. If now $K(x, \xi; s, \sigma) \geqq 0$ denotes the Green's function of the problem

$$\Delta y - y \exp w = 0, \qquad y = 0 \text{ on } C,$$

then it may be shown that $y(x, \xi)$ is a solution of the nonlinear integral equation

$$y(x, \xi) + \iint_B K(x, \xi; s, \sigma)[\exp (y) - y] \exp (w)\, ds\, d\sigma = 0.$$

iii. In the discussion of certain wave phenomena L. Lichtenstein [3] was led to the study of integral equations of the form

$$Ly(x) - \sum_{i=1}^{\infty} \int_a^b \cdots \int_a^b K_i(x, s_1 \cdots s_i)y(s_1) \cdots y(s_i)\, ds_1 \cdots ds_i = 0,$$

with $L$ denoting a linear operator, and the $K_i$ symmetric and continuous functions in the $(n + 1)$ dimensional unit cube.

Physical situations leading to similar nonlinear integral equations have also been described by Schmeidler [4].

The solution of such integral equations may be accomplished in general by a variety of methods, with the choice of the most appropriate one naturally depending upon the form and type of the integral equations of interest. Such methods may employ closed analytical expressions, power series expansions, the replacement of integrations by summations, successive approximations, variational methods, etc. [5].

The approach to be used here will be the iterative one, to be applied to integral equations of a generalized Lichtenstein type:

$$y(x) - \sum_{i=1}^{r} \int_a^b \cdots \int_a^b K_i(x, s_1 \cdots s_i) F_i(s_1, \cdots s_i) y(s_1) \cdots y(s_i)) \, ds_1 \cdots ds_i = 0. \quad (1)$$

Integral equations containing an "inhomogeneous" term, e.g., of the form

$$y^*(x) - K_0(x) - \int K_1 F_1^*(y^*) - \iint K_2 F_2^*(y^*) = 0$$

can obviously be reduced to the form (1) by the substitution

$$y = y^* - K_0 .$$

For kernels $K_i$ whose squares are integrable, and functions $F_i$ that have bounded derivatives the existence and uniqueness of a solution will be proved, and an upper bound for $\max \mid y \mid$ will be given. The proof is based on the method of successive iterations, a method which also easily provides an estimate of the error inherent in the $n$th approximation. The results will then be generalized to systems of equations of type (1), and, finally, an example will be given.

**2. Assumptions and Notations.** For the sake of simplicity, we shall take $r = 2$ in eq. (1), i.e. consider in the interval $I: a \leqq x, s, t \leqq b$ the equation

$$y(x) - \int K_1(x, s) F_1(s, y(s)) \, ds - \iint K_2(x, s, t) F_2(s, t, y(s), y(t)) \, ds \, dt = 0. \quad (2)$$

The kernels $K_1$ and $K_2$ will be assumed to have integrable squares, so that $\int K_1^2(x, s) \, ds$, $\iint K_2^2(x, s, t) \, ds \, dt$ exist and are continuous functions of $x \varepsilon I$. For the maximum of a continuous function $f(s)$ defined in $I$ we shall employ the notation

$$\| f \| = \max_{x \varepsilon I} \mid f(x) \mid .$$

Then

$$K^2 = \left\| \int K_1^2 \, ds + \iint K_2^2 \, ds \, dt \right\| \quad \text{certainly exists.}$$

The continuous functions $F_1(s, u)$, $F_2(s, t, u, v)$, defined for $s, t \varepsilon I$; $u, v \varepsilon J$,

where $J$ denotes an interval $\langle -c, c\rangle$, are assumed to satisfy conditions of the form

$$| F_1(s, u_1) - F_1(s, u_2) | \leqq L_1 | u_1 - u_2 | \tag{3}$$

$$| F_2(s, t, u_1, v_1) - F_2(s, t, u_2, v_2) | \leqq L_2(| u_1 - u_2 | + | v_1 - v_2 |),$$

where $L_1$, $L_2$ are positive numbers independent of $s, t, u, v$. The quantities

$$M_1^2 = \| F_1^2(s, 0) \|, \qquad M_2^2 = \| F_2^2(s, t, 0, 0) \|$$

then also exist.

Relations (3) certainly hold if, for example, $F_1$ and $F_2$ have continuous partial derivatives $\partial F_1/\partial u$, $\partial F_2/\partial u$, $\partial F_2/\partial v$ for arguments restricted, respectively, to the intervals $I$, $J$.

Further, we put

$$A^2 = b - a, \qquad \alpha = 1 \quad \text{or } 2 \text{ for } \quad A \leqq 1 \text{ or } A > 1.$$

Other quantities to be used later on are

$$L^2 = L_1^2 + L_2^2, M^2 = M_1^2 + M_2^2, \qquad A^\alpha KM = E, \qquad A^\alpha KL = D.$$

In the following extensive use will be made of Schwarz's inequality

$$\left| \int f(x)g(x)\, dx \right| \leqq \left( \int f^2\, dx \right)^{1/2} \left( \int g^2\, dx \right)^{1/2}. \tag{4}$$

This relation is applicable also to multiple integrals and sums. If we put

$$\left| \int f \right| = I^2(f^{1/2}),$$

then eq. (4) may be expressed in the form

$$I^2(f^{1/2}g^{1/2}) \leqq I(f)I(g). \tag{5}$$

**3. Lemma.** We are now in a position to state the following LEMMA:

Let $y \varepsilon J$ be a solution of eq. (2), and let the $F_i$ satisfy conditions of type (3). Then $\| y \| (1 - 2D) \leqq E$.

Proof. It follows from eq. (2) that

$$| y | \leqq \left| \int K_1 F_1 \right| + \left| \iint K_2 F_2 \right|.$$

In this inequality the right side is, by (5), identically in $x$,

$$\leqq I(K_1)I(F_1) + I(K_2)I(F_2).$$

From Schwarz's inequality for sums it may now be deduced that

$$| y | \leqq [I^2(K_1) + I^2(K_2)]^{1/2}[I^2(F_1) + I^2(F_2)]^{1/2}$$

$$\leqq K[I^2(F_1) + I^2(F_2)]^{1/2}.$$

Since, by assumption, $y \varepsilon J$, we may make use of relationships (3), getting

$$I^2(F_2) \leqq \iint [\,|F_2(s,t,0,0)| + L_2(|y(s)| + |y(t)|)]^2\,ds\,dt$$
$$\leqq A^{2\alpha}(M_2 + 2L_2\|y\|)^2.$$

Similarly,

$$I^2(F_1) \leqq A^{2\alpha}(M_1 + 2L_1\|y\|)^2.$$

By addition, then,

$$I^2(F_1) + I^2(F_2) \leqq A^{2\alpha}(M + 2L\|y\|)^2.$$

Thus

$$\|y\| \leqq KA^{\alpha}(M + 2L\|y\|) = E + 2D\|y\|,$$

which proves the lemma.

**4. Existence Theorem.** For functions $F_1$, $F_2$ satisfying conditions (3) with

$$0 \leqq E/(1 - 2D) \leqq c \tag{6}$$

eq. (2) has exactly one solution.

To prove this theorem we use successive iteration, starting with an arbitrary continuous function $y_0 \varepsilon J$, and defining a sequence $\{y_n\}$ of continuous functions by

$$y_{n+1}(x) = \int K_1(x,s)F_1(s, y_n(s))\,ds + \iint K_2(x,s,t)F_2(s,t,y_n(s),y_n(t))\,ds\,dt, \tag{7}$$
$$n = 0, 1, 2, \cdots$$

First we show by induction that $y_n \varepsilon J$. Let, then, $y_{n-1}$ have this property. By definition (7), following the same reasoning that was used to prove the lemma, there is obtained.

$$\|y_n\| \leqq E + 2D\|y_{n-1}\|.$$

From the inductive assumption and condition (6) it may now be inferred that

$$\|y_n\| \leqq E + 2Dc \leqq c,$$

i.e. $y_n \varepsilon J$. Next, the sequence is easily shown to be uniformly convergent, as follows: From

$$|y_{n+1} - y_n| \leqq \left| \int K_1(F_1(y_n) - F_1(y_{n-1}))\,ds \right|$$
$$+ \left| \iint K_2(F_2(y_n) - F_2(y_{n-1}))\,ds\,dt \right|$$

it follows, by means of Schwarz's inequality, that

$$| y_{n+1} - y_n | \leqq K[I^2(F_1(y_n) - F_1(y_{n-1}) + I^2(F_2(y_n) - F_2(y_{n-1}))]^{1/2}.$$

Taking advantage of relations (3) the right side above becomes

$$\leqq 2KA^aL \parallel y_n - y_{n-1} \parallel,$$

so that

$$\parallel y_{n+1} - y_n \parallel \leqq 2D \parallel y_n - y_{n-1} \parallel, \tag{8}$$

or, by induction,

$$\parallel y_{n+1} - y_n \parallel \leqq \parallel y_1 - y_0 \parallel (2D)^n, \qquad n \geqq 0.$$

Since $1 - 2D > 0$, $\sum_{n=0}^{\infty} (2D)^n$ is convergent, and it follows that the partial sums $y_n$ of $y_0 + \sum_{n=1}^{\infty} (y_n - y_{n-1})$ are absolutely and uniformly convergent.

The continuous limit function $y$ of the $\{y_n\}$ is easily shown to be a solution of eq. (2). Putting

$$\int K_1 F_1(y)\, ds + \iint K_2 F_2(y)\, ds\, dt = y^*(x)$$

we find, namely, that

$$| y_{n+1} - y^* | \leqq K[I^2(F_1(y_n) - F_1(y)) + I^2(F_2(y_n) - F_2(y))]^{1/2}.$$

But the continuity of the $F_i$ is conjunction with the uniform convergence of $y_n$ towards $y$ suffice to make the difference $F_i(y_n) - F_i(y)$ arbitrarily small by taking $n$ sufficiently large. Thus also $\lim_n y_n = y^*$ uniformly in $x$. It follows that $y^* = y$, i.e. that $y$ is a solution of eq. (2).

Finally, to show that $y$ is the only solution of eq. (2) under the conditions stated above let it be assumed that there exists another solution $Y(x)$. Clearly

$$\parallel Y - y \parallel \leqq \parallel y_{n+1} - Y \parallel + \parallel y_{n+1} - y \parallel,$$

and $\parallel y_{n+1} - y \parallel < \varepsilon/2$ for $n > N_1$. The reasoning applied above to establish the relationship (8) now leads to

$$\parallel y_{n+1} - Y \parallel \leqq 2D \parallel y_n - Y \parallel,$$

or

$$\parallel y_{n+1} - Y \parallel \leqq (2D)^n \parallel y_0 - Y \parallel,$$

whence

$$\parallel y_{n+1} - Y \parallel < \varepsilon/2 \text{ for } n > N_2.$$

Consequently

$$\parallel Y - y \parallel < \varepsilon.$$

Due to the arbitrarinesss of $\varepsilon$ and the meaning of $\parallel Y - y \parallel$ it is then necessary that $Y(x) - y(x) \equiv 0$.

By the lemma which is applicable here, and by condition (6),

$$\| y \| \leqq E/(1 - 2D) \leqq c.$$

For increasing values of $K$, $L$, $M$, therefore, the upper bound $E/(1 - 2D)$ of $\| y \|$ also increases.

An estimate of the error $\| y - y_n \|$ of the $n^{\text{th}}$ approximation $y_n$ is easily obtained. From

$$\| y - y_n \| \leqq \| y_{n+1} - y_n \| + \| y - y_{n+1} \|$$

it follows that

$$\| y - y_n \| \leqq \sum_{i=n}^{n+m} \| y_{i+1} - y_i \| + \| y - y_{n+m+1} \|.$$

For fixed $n$, $m \to \infty$, then, since $y_m \to y$ uniformly,

$$\| y - y_n \| \leqq \sum_{i=n}^{\infty} \| y_{i+1} - y_i \| = \sum_{i=0}^{\infty} \| y_{i+1} - y_i \| - \sum_{i=0}^{n-1} \| y_{i+1} - y_i \|.$$

Thus

$$\| y - y_n \| \leqq \| y_1 - y_0 \| (2D)^n/(1 - 2D) \quad \text{for} \quad n \geqq 0. \tag{9}$$

**5. Generalizations.** Equation (2) may be considered to be a special case of a more general system

$$y_i(x) - \sum_{j=1}^{q} \int \cdots \int K_{ij}(x, s_1, \cdots, s_j) F_{ij}(s_1, \cdots s_j, y_1(s_1), \cdots y_1(s_j), \cdots, y_p(s_1)), \cdots, y_p(s_j))\, ds_1 \cdots ds_j = 0,$$

where $i = 1, 2, \cdots, p$, $p \geqq 1$, $q \geqq 1$. To such systems the results of the previous sections may be transferred without difficulty; it is necessary only to replace the quantities $\alpha$, $E$, $D$, $\| y \|$, etc. by

$$\alpha = 1 \quad \text{or } q \text{ for} \quad A \leqq 1 \quad \text{or} \quad A > 1,$$

$$\| \textstyle\sum_j I^2(K_{ij}) \| = K_i^2,$$

$$\textstyle\sum_j M_{ij}^2 = M_i^2, \qquad A^\alpha \sum_i K_i M_i = E, \qquad A^\alpha \sum_i K_i L_i = D,$$

$$\| y \| = \| \textstyle\sum_i y_i(x) \|.$$

Finally, the important subsidiary condition $1 - 2D > 0$ must be replaced by $1 - qD > 0$.

For $p = q = 1$, i.e., for the equation

$$y(x) - \int K(x, s) F(s, y(s))\, ds = 0, \tag{10}$$

this condition $1 - qD > 0$ becomes $0 < AKL < 1$ or

$$0 < A^2K^2L^2 < 1. \tag{11}$$

Let us ascertain its meaning for the case of the linear integral equation of the second kind

$$y(x) - \lambda \int_a^b K(x, s)y(s)\, ds = f(x). \tag{12}$$

This type of equation is obtained from eq. (10) by putting

$$F(s, u) = g(s) + \lambda u$$

$$f(x) = \int K(x, s)g(s)\, ds.$$

Since in this case $L = \| \partial F/\partial u \| = |\lambda|$, inequality (11) indicates that

$$\lambda^2 \iint K^2(s, t)\, ds\, dt \leqq L^2A^2K^2 < 1.$$

Now for the characteristic value of smallest absolute amount of the symmetric kernel there exists the relation [6]

$$\lambda_1^2 \geqq \left(\iint K^2\, ds\, dt\right)^{-1}.$$

Consequently, condition (11) becomes

$$|\lambda| < |\lambda_1|,$$

which is the well-known sufficient condition for the existence of a unique solution of eq. (12) for arbitrary $f(x)$.

**6. Applications.** There are many practical cases where the kernels may be expressed in bilinear forms

$$\begin{aligned} K_1(x, s) &= \textstyle\sum_i \phi_{1i}(x)\Psi_{1i}(s) \\ K_2(x, s, t) &= \textstyle\sum_j \phi_{2j}(x)\Psi_{2j}(s, t) \end{aligned} \tag{13}$$

the $\phi_{1i}$ being linearly independent of the $\phi_{2j}$, and the summation to be extended over a finite or infinite number of values $i, j$. In such cases the iterative procedure (7) clearly becomes

$$\begin{aligned} y_{n+1}(x) &= \sum_i \phi_{1i}(x)p_{n+1,i} + \sum_j \phi_{2j}(x)q_{n+1,j} \\ p_{n+1,i} &= \int \psi_{1i}(s)F_1\left(s, \sum_i \phi_{1i}(s)p_{ni} + \sum_j \phi_{2j}(s)q_{nj}\right) ds \\ q_{n+1,i} &= \iint \psi_{2j}\, F_2\left(s, t, \sum_i (s) + \sum_j (s), \sum_i (t) + \sum_j (t)\right) ds\, dt. \end{aligned} \tag{14}$$

The iterative procedure for the functions $y_n(x)$ is thus reduced to a method of successive approximation for the constants $p_{ni}$, $q_{nj}$:

$$p_{n+1,i} = \Phi_1(p_{ni}, q_{nj}), \qquad q_{n+1,j} = \Phi_2(p_{ni}, q_{nj}). \tag{15}$$

In the limit, then,

$$y(x) = \sum_i \phi_{1i}(x)p_i + \sum_j \phi_{2j}(x)q_j, \tag{16}$$

where the $p_i$, $q_j$ satisfy the functional equations

$$p_i = \Phi_1(p_i, q_j), \qquad q_j = \Phi_2(p_i, q_j) \quad \text{for all } i, j. \tag{17}$$

Whenever the functional equations (17) are directly solvable the solution (16) may be obtained immediately. Otherwise use must be made of the successive approximations (15).

Let us now apply the foregoing deductions to the following case:

$$K_1(x, s) = x(1 - s), \qquad I: \langle 0, 1\rangle$$

$$K_2(x, s, t) = \exp(x - s - t)$$

$$F_1(s, u) = \lambda(u + 1) \qquad J: \langle -1, 1\rangle$$

$$F_2(s, t, u, v) = \lambda uv.$$

The equation to be solved, then, is

$$y(x) - \lambda\left\{\int_0^1 x(1 - s)[y(s) + 1]\,ds + \int_0^1\int_0^1 \exp(x - s - t)y(s)y(t)\,ds\,dt\right\} = 0. \tag{18}$$

For the convergence of the iterative procedure discussed above it is sufficient that

$$\theta \leqq E/(1 - 2D) \leqq 1. \tag{6}$$

Now

$$A^a = 1, \qquad K = [(1/3) + (e^2 - 1)^2/4e^2]^{1/2} \approx 1.31$$

$$L = 2^{1/2}\,|\lambda|, \qquad M = |\lambda|.$$

Condition (6) is satisfied, then, for

$$\lambda \leqq \lambda_0 \equiv [K(1 + 2\sqrt{2})]^{-1} \approx 0.2. \tag{19}$$

Further, the lemma indicates that for the solution $y(x)$,

$$\|\, y(x) \,\| \leqq K\,|\lambda|/(1 - 2K\,|\lambda|\,\sqrt{2}).$$

Since in this example $i = j = 1$, the iterative process for eq. (18) may be expressed in the form

$$y_{n+1}(x) = \lambda(p_{n+1}x + q_{n+1}\exp(x))$$

with

$$p_{n+1} = \int_0^1 (1 - s)[\lambda(p_n s + q_n \exp(s) + 1]\,ds$$

$$q_{n+1} = \lambda^2\left[\int_0^1 (p_n s \exp(-s) + q_n)\,ds\right]^2.$$

According to the existence and uniqueness theorem $p = \lim_n p_n$, $q = \lim_n q_n$ exist. These limits satisfy a system of equations corresponding to equations (17), in this case

$$p(1 - \tfrac{1}{6}\lambda) - q\lambda(e - 2) = \tfrac{1}{2}, \qquad \lambda^2[(1 - 2/e)p + q]^2 = q.$$

Eliminating $p$ now there is obtained for $q$ the equation

$$\lambda^2(\alpha p + \beta)^2 = \gamma q \tag{20}$$

with

$$\alpha = 6\lambda(e - 2)^2 + (6 - \lambda)e, \qquad \beta = 3(e - 2), \qquad \gamma = e^2(6 - \lambda)^2.$$

Consequently,

$$q = \{(\gamma - 2\alpha\beta\lambda^2) \pm [\gamma(\gamma - 4\alpha\beta\lambda^2)]^{1/2}\}/2\alpha^2\lambda^2. \tag{21}$$

Of the two possible solutions indicated here the one corresponding to the positive sign becomes unbounded as $\lambda$ approaches zero. Since it is desired to keep the solutions of norm not exceeding unity the positive sign in eq. (21) must be discarded.

Suppose we wish to determine the solution of eq. (18) for $\lambda = 0.2$. For the initial approximation we take

$$p_0 = q_0 = 1,$$

i.e.

$$y_0(x) = 0.2(x + \exp(x)).$$

The resulting approximations

$$p_{n+1} = 0.2[\tfrac{1}{6}p_n + q_n(e - 2)] + \tfrac{1}{2}, \qquad q_{n+1} = 0.04[(e - 2)p_n/e + q_n]^2$$

are shown in Table I.

TABLE I.

*Successive Approximation*

| $n$ | $p_n$ | $q_n$ |
|---|---|---|
| 0 | 1.0000 | 1.0000 |
| 1 | 0.6770 | 0.0639 |
| 2 | 0.5317 | 0.0024 |
| 3 | 0.5181 | 0.0008 |
| 4 | 0.5174 | 0.0008 |

The exact solution, as determined from eq. (21), is

$$y(x) = 0.2(0.51735x + 0.00075 \exp(x)).$$

While there is always convergence to the solution for $y_0 \varepsilon J$, $|\lambda| \leqq \lambda_0$, the choice of an initial approximation $y_0$ outside of $J$, or a value of $|\lambda| > \lambda_0$ may

lead to divergence. This is the case, for example, for $p_0 = q_0 = 25$, $\lambda = 0.2$, and $p_0 = q_0 = 1$, $\lambda = 1$.

It is worthwhile mentioning that iterative procedures such as those exemplified by eqs. (15) are eminently suited for treatment on high speed computing machinery. The required quadratures may be performed by any suitable standard methods, or by means of the formulas outlined in [7].

## REFERENCES

1. Nyström, E. J. Zur numerischen Lösung von Randwertaufgaben bei gewöhnlichen Differentialgleichungen, Acta Math. **76**, 157–184, 1945.
2. Bieberbach, L. Göttinger Nachrichten 1912, 599–602.
3. Lichtenstein, L. Vorlesungen über einige Klassen nichtlinearer Integralgleichungen, Berlin, 1931.
4. Schmeidler, W. Algebraische Integralgleichungen, Math. Nachr. **8**, 31–40, 1952.
5. Collatz, L. Numerische Behandlung von Differentialgleichungen, Berlin, 1951, p. 370, etc.
6. Schmidt, E. Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener, Math. Ann. **63**, 455.
7. Lotkin, M. A New Integration Procedure, Journal Math. Phys. **32**, 171–179, 1953.

RCA Service Company
Patrick Air Force Base, Florida.

(Received December, 1953)

# A SIMPLE METHOD FOR SUMMING CERTAIN SLOWLY CONVERGENT SERIES

By Herbert E. Salzer

Although there has been a fairly substantial literature as noted in the bibliography on various devices and transformations to speed the convergence of slowly convergent series (or sequences), the writer has not encountered the following method of summation which would seem to be one of the most obvious and convenient, and whose application may be widespread since it involves a purely numerical device which is employed without any special analytic work upon the series. The method is to consider $S_j$, the $j^{\text{th}}$ term of a sequence (which is usually the sum of the first $j$ terms of a series of the requisite type), as the tabulated value of a certain function of $x$, $S(x)$, at $x = j$, from which one would like to calculate $S(\infty)$ by some kind of extrapolation. Suppose that the last term of the sequence is for $j = n$. Instead of trying to extrapolate directly from $x = n$ to $x = \infty$, it is convenient to consider the new variable $y = 1/x$ and to denote $S(x)$ by $\bar{S}(y) \equiv \bar{S}(1/x)$. Now the problem is to extrapolate for $\bar{S}(y)$ at $y = 0$ from its values at $y = 1/n, 1/(n-1), 1/(n-2)$, etc. where we do not go all the way back to the first term corresponding to $y = 1/1$, but take the last $m$ terms of the sequence $S_j$, namely $S_n, S_{n-1}, \cdots, S_{n-m+1}$. The extrapolation formula is obtained from the $m$-point Lagrangian interpolation polynomial for $\bar{S}(y)$ at $y = 1/n, 1/(n-1), \cdots, 1/(n-m+1)$, by setting $y$ equal to 0. Thus to calculate the limit $S$ of a sequence $S_1, S_2, \cdots, S_j$, which is given up to $j = n$, using the $m$-point extrapolation formula, one merely multiplies each of the last $m$ terms, $S_{n-i}$, $i = 0, 1, \cdots, m-1$, by the corresponding extrapolation coefficient $A^{(m)}_{n,n-i}$, and sums, thus,

$$S \sim \sum_{i=0}^{m-1} A^{(m)}_{n,n-i} S_{n-i}. \tag{1}$$

Table I gives the exact values of the coefficients $A^{(m)}_{n,n-i}$, $i = 0, 1, 2, \cdots, m-1$, for $m = 4, 7$ and 11 (i.e. 3rd, 6th and 10th degree interpolation polynomial approximation to $\bar{S}(y)$, for $n = 5, 10, 15$ and 20 (there being, of course, no $m = 7$ and 11 for $n = 5$ and no $m = 11$ for $n = 10$). For convenience in machine computation, the $A^{(m)}_{n,n-i}$ is expressed as $B^{(m)}_{n,n-i}/D^{(m)}_n$, where for each $m$ and $n$, $D^{(m)}_n$ is the least integer which will allow every $B^{(m)}_{n,n-i}$ to be integral. Thus (1) is employed in the form

$$S \sim (1/D^{(m)}_n) \sum_{i=0}^{m-1} B^{(m)}_{n,n-i} S_{n-i}. \tag{2}$$

The choice of $m$ and $n$ in Table I is intended to be fairly adequate for the very many instances where (2) is applicable. The user can very easily compute

TABLE I

Table of $A_{n,n-i}^{(m)} = B_{n,n-i}^{(m)}/D_n^{(m)}$

| $n = 5$ | $n = 10$ | $n = 15$ | $n = 20$ |
|---|---|---|---|
| 4-point | 4-point | 4-point | 4-point |
| $B_{5,2}^{(4)} = -8$ | $B_{10,7}^{(4)} = -343$ | $B_{15,12}^{(4)} = -576$ | $B_{20,17}^{(4)} = -4913$ |
| $B_{5,3}^{(4)} = 81$ | $B_{10,8}^{(4)} = 1536$ | $B_{15,13}^{(4)} = 2197$ | $B_{20,18}^{(4)} = 17496$ |
| $B_{5,4}^{(4)} = -192$ | $B_{10,9}^{(4)} = -2187$ | $B_{15,14}^{(4)} = -2744$ | $B_{20,19}^{(4)} = -20577$ |
| $B_{5,5}^{(4)} = 125$ | $B_{10,10}^{(4)} = 1000$ | $B_{15,15}^{(4)} = 1125$ | $B_{20,20}^{(4)} = 8000$ |
| $D_5^{(4)} = 6$ | $D_{10}^{(4)} = 6$ | $D_{15}^{(4)} = 2$ | $D_{20}^{(4)} = 6$ |
| | 7-point | 7-point | 7-point |
| | $B_{10,4}^{(7)} = 2048$ | $B_{15,9}^{(7)} = 1\ 77147$ | $B_{20,14}^{(7)} = 37\ 64768$ |
| | $B_{10,5}^{(7)} = -46875$ | $B_{15,10}^{(7)} = -20\ 00000$ | $B_{20,15}^{(7)} = -341\ 71875$ |
| | $B_{10,6}^{(7)} = 3\ 49920$ | $B_{15,11}^{(7)} = 88\ 57805$ | $B_{20,16}^{(7)} = 1258\ 29120$ |
| | $B_{10,7}^{(7)} = -11\ 76490$ | $B_{15,12}^{(7)} = -199\ 06560$ | $B_{20,17}^{(7)} = -2413\ 75690$ |
| | $B_{10,8}^{(7)} = 19\ 66080$ | $B_{15,13}^{(7)} = 241\ 34045$ | $B_{20,18}^{(7)} = 2550\ 91680$ |
| | $B_{10,9}^{(7)} = -15\ 94323$ | $B_{15,14}^{(7)} = -150\ 59072$ | $B_{20,19}^{(7)} = -1411\ 37643$ |
| | $B_{10,10}^{(7)} = 5\ 00000$ | $B_{15,15}^{(7)} = 37\ 96875$ | $B_{20,20}^{(7)} = 320\ 00000$ |
| | $D_{10}^{(7)} = 360$ | $D_{15}^{(7)} = 240$ | $D_{20}^{(7)} = 360$ |
| | | 11-point | 11-point |
| | | $B_{15,5}^{(11)} = 19\ 53125$ | $B_{20,10}^{(11)} = 10000\ 00000$ |
| | | $B_{15,6}^{(11)} = -1209\ 32352$ | $B_{20,11}^{(11)} = -2\ 59374\ 24601$ |
| | | $B_{15,7}^{(11)} = 25422\ 77241$ | $B_{20,12}^{(11)} = 27\ 86281\ 39008$ |
| | | $B_{15,8}^{(11)} = -2\ 57698\ 03776$ | $B_{20,13}^{(11)} = -165\ 43019\ 02188$ |
| | | $B_{15,9}^{(11)} = 14\ 64449\ 44842$ | $B_{20,14}^{(11)} = 607\ 43477\ 54496$ |
| | | $B_{15,10}^{(11)} = -50\ 40000\ 00000$ | $B_{20,15}^{(11)} = -1453\ 15898\ 43750$ |
| | | $B_{15,11}^{(11)} = 108\ 93718\ 33242$ | $B_{20,16}^{(11)} = 2308\ 97441\ 83296$ |
| | | $B_{15,12}^{(11)} = -148\ 60167\ 41376$ | $B_{20,17}^{(11)} = -2419\ 19268\ 05388$ |
| | | $B_{15,13}^{(11)} = 124\ 07264\ 26641$ | $B_{20,18}^{(11)} = 1606\ 71025\ 19808$ |
| | | $B_{15,14}^{(11)} = -57\ 85093\ 09952$ | $B_{20,19}^{(11)} = -613\ 10662\ 57801$ |
| | | $B_{15,15}^{(11)} = 11\ 53300\ 78125$ | $B_{20,20}^{(11)} = 102\ 40000\ 00000$ |
| | | $D_{15}^{(11)} = 725760$ | $D_{20}^{(11)} = 362880$ |

the coefficients $A^{(m)}_{n,n-i}$, $i = 0, 1, 2, \cdots, m-1$, for other values of $m$ and $n$ from the simple formula

$$A^{(m)}_{n,n-i} = \frac{(-1)^i(n-i)^{m-1}}{(m-1)!}\binom{m-1}{i}, \tag{3}$$

where $\binom{m-1}{i}$ is a binomial coefficient.

In using this table for some particular $m$ and $n$ the computer must have enough extra significant figures in the $S_j$ terms in order to obtain the full accuracy of (2). For slowly convergent sequences, since the $S_j$ terms are nearly equal, when $k$-place accuracy is expected in the answer for $S$, it is best to carry $k + l + 2$ places in the $S_j$, where $l$ is the number of digits in the $A^{(m)}_{n,n-i}$ having the largest absolute value.

This method of summing was tested upon four different slowly convergent sequences for various values of $m$ and $n$. In every case tested, whether for $n = 10$, 15 or 20, at the outset the agreement of $S_n$ with $S$ was never beyond the first decimal place (in fact, the series for $\pi^2/6$ does not give even 1 decimal place accuracy for $n = 10$). But the use of (2) upon those $S_j$ of only 1-place accuracy, gave as high as 15-place accuracy in certain cases.

In the series for $\pi/4$ and log 2, which were alternating in sign, the method failed when used upon a number of consecutive partial sums. However, by separating the partial sums of an odd number of terms from the partial sums of an even number of terms and by applying (2) to either $S_j \equiv \sum_{r=1}^{2j-1} u_r$ or $S_j \equiv \sum_{r=1}^{2j} u_r$, where $u_r = (-1)^{r-1}/(2r-1)$ for $\pi/4$ and $u_r = (-1)^{r-1}/r$ for log 2, one obtained accuracy comparable to that in the other examples.

Following are the details of the tests.

a) The sequence for Euler's constant $\gamma \equiv \lim_{j\to\infty}\{\sum_{r=1}^{j} r^{-1} - \log j\} = 0.57721\ 56649\ 01532\ 86061$ to twenty places, was tested by considering $S_j = \sum_{r=1}^{j} r^{-1} - \log j$ up to $j = 20$. The use of (2) to calculate $\gamma$, for $n = 20$ gave this accuracy: for $m = 4$, to within 1 unit in the 7th decimal place, for $m = 7$ to within $\frac{1}{2}$ unit in the 11th decimal place, and for $m = 11$, to within $1\frac{1}{2}$ units in the 15th decimal place.

b) The sequence for $\pi^2/6 = \sum_{r=1}^{\infty} r^{-2} = 1.64493\ 40668\ 48$ to twelve places, was tested by considering $S_j' = \sum_{r=1}^{j} r^{-2}$ up to $j = 10$. The use of (2) for $n = 10$ gave for $m = 4$, accuracy to within $\frac{1}{2}$ unit in the 5th decimal place (6th significant figure) and for $m = 7$ accuracy to within $1\frac{1}{2}$ units in the 8th decimal place (9th significant figure).

c) The sequence for $\pi/4 = \sum_{r=1}^{\infty}(-1)^{r-1}/(2r-1) = 0.78539\ 81633\ 97$ to twelve places, was tested for odd sums $S_j = \sum_{r=1}^{2j-1}(-1)^{r-1}/(2r-1)$ and even sums $S_j = \sum_{r=1}^{2j}(-1)^{r-1}/(2r-1)$, up to $j = 10$. The use of (2) for $n = 10$ gave for $m = 4$ accuracy to within $\frac{2}{3}$ unit in the 6th decimal place for odd sums and to within $\frac{1}{4}$ unit in the 6th decimal place for even sums, and for $m = 7$ accuracy to within $1\frac{1}{2}$ units in the 9th decimal place for both odd and even sums.

d) The most comprehensive of the tests was made upon the sequence for

$\log 2 = \sum_{r=1}^{\infty} (-1)^{r-1}/r = 0.69314\ 71805\ 59945\ 30942$ to twenty places. Again both odd sums $S_j = \sum_{r=1}^{2j-1} (-1)^{r-1}/r$ and even sums $S_j = \sum_{r=1}^{2j} (-1)^{r-1}/r$ were considered as far as $j = 20$. The use of (2) for $n = 10$ gave for $m = 4$ accuracy to within $1\frac{1}{2}$ units in the 6th decimal place and for $m = 7$ accuracy to within $\frac{1}{2}$ unit in the 8th decimal place, for both odd and even sums. The use of (2) for $n = 15$ gave for $m = 4$ accuracy to within $\frac{1}{4}$ unit in the 6th decimal place, for $m = 7$ accuracy to within 1 unit in the 10th decimal place, and for $m = 11$ accuracy to within $1\frac{1}{2}$ units in the 13th decimal place, for both odd and even sums. The use of (2) for $n = 20$ gave for $m = 4$ accuracy to within 1 unit in the 7th decimal place and for $m = 7$ accuracy to within $\frac{1}{2}$ unit of the 11th decimal place, for both odd and even sums, and for $m = 11$ accuracy to within 2 units in the 15th decimal place for odd sums and to within $1\frac{1}{2}$ units in the 15th decimal place for even sums.

In both c) and d) the reader will note the very close agreement between the answers obtained from (2) using the odd sums and those obtained using the even sums.

Thus far the only type of sequence $S_j$ upon which this method has been adequately tested is the very slowly convergent monotonic type. This method might be tried upon series whose terms alternate or vary in sign, or whose terms may even oscillate in value following some kind of predictable pattern, provided that it is possible to regroup them (without altering $S$) into a slowly convergent monotonic sequence.

## BIBLIOGRAPHY

[1] W. G. Bickley and J. C. P. Miller, "The Numerical Summation of Slowly Convergent Series of Positive Terms," *Phil. Mag.*, 7th ser., **22**, pp. 754–767, (1936).

[2] J. W. Bradshaw, "Continued Fractions and Modified Continued Fractions for Certain Series," *Amer. Math. Monthly*, **45**, no. 6, pp. 352–362, (1938).

[3] J. W. Bradshaw, "Modified Series," *Amer. Math. Monthly*, **46**, no. 8, pp. 486–492, (1939).

[4] T. Fort, *Finite Differences*, Oxford, Clarendon Press, pp. 18–23, (1948).

[5] K. Knopp, *Theory and Application of Infinite Series*, 2nd English ed., Blackie and Son Ltd., London and Glasgow, Chapter VIII, pp. 230–273, (1951).

[6] S. Lubkin, "A Method of Summing Infinite Series," *Jour. Res. Nat. Bur. Stands.*, **48**, no. 3, pp. 228–254, RP 2310, (1952).

[7] J. Barkley Rosser, "Transformations to Speed the Convergence of Series," *Jour. Res. Nat. Bur. Stands.*, **46**, no. 1, pp. 56–64, RP 2175, (1951).

[8] J. A. Shohat, "On a Certain Transformation of Infinite Series," *Amer. Math. Monthly*, **40**, no. 4, pp. 226–229, (1933).

[9] O. Szász, "Summation of Slowly Convergent Series," *Jour. Math. and Physics*, **28**, no. 4, pp. 272–279, (1950).

[10] E. Whittaker and G. Robinson, *The Calculus of Observations*, 4th ed., Blackie and Son Ltd., London and Glasgow, pp. 368–369, (1946).

Further references to the literature are given in [2], [5], [6], [7], and [10] above.

Diamond Ordnance Fuze Laboratories
Washington, D.C.

(Received January 8, 1954)

# TABLES OF SELF-EQUILIBRATING FUNCTIONS

By G. Horvay and J. S. Born

It is shown in Reference 1[1] that the problems of finding stress distributions in semi-infinite strips due (a) to prescribed self-equilibrating normal tractions $\bar{\sigma}_x$, and (b) to prescribed self-equilibrating shear tractions $\bar{\tau}$, applied at the left edge as in Fig. 1 (bar above a quantity denotes edge value for $x = 0$), permit approximate solutions as follows.

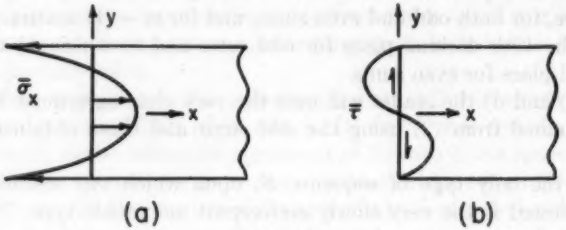



Fig. 1

We expand the known boundary values $\bar{\varphi}(y)$ (in Case (a), $\bar{\varphi}_{,x}(y)$ in Case (b)) of the unknown stress function $\varphi(x, y)$ into polynomials $f_n(y)$:

*Case a*

$$\int_A^y d\eta \int_B^\eta \bar{\sigma}_x(\eta_1)\, d\eta_1 = \bar{\varphi}(y) = \langle \bar{\varphi}, f_2 \rangle f_2 + \langle \bar{\varphi}, f_3 \rangle f_3 + \cdots \tag{1}$$

*Case b*

$$-\int_A^y \bar{\tau}(\eta)\, d\eta = \bar{\varphi}_{,x}(y) = \langle \bar{\varphi}_{,x}, f_2 \rangle f_2 + \langle \bar{\varphi}_{,x}, f_3 \rangle f_3 + \cdots \tag{2}$$

where

$$\langle \bar{\varphi}, f \rangle \equiv \int_{y=-1}^{y=+1} \bar{\varphi} f\, dy \tag{3}$$

and

$B = 0$, $A = -1$ for $\bar{\sigma}_x$ = even function of $y$, $\bar{\tau}$ = odd function of $y$

$B = -1$, $A = 0$ for $\bar{\sigma}_x$ = odd function of $y$, $\bar{\tau}$ = even function of $y$.

Here $f_n(y)$ constitute an orthonormal set of functions, listed in Table Ia for $n = 2$ to 9, which obey the differential equation

$$\begin{aligned}(1 - y^2)^2 f^{IV} - 10y(1 - y^2) f''' + (\lambda - 24)(1 - y^2) f'' \\ + (12 - 4\lambda) y f' - 2\lambda f = 0\end{aligned} \tag{4}$$

$$\lambda = (n + 2)(n + 3)$$

[1] G. Horvay, The End Problem of Rectangular Strips, J. Appl. Mech. *20*, 87 (1953).

## TABLE IA

*Self-equilibrating polynomials, $f_n = N_n \chi_n$*

| $n$ | $N_n$ | $\chi_n$ |
|---|---|---|
| 2 | $\frac{3}{16}\sqrt{5\cdot 7} = 1.1092650$ | $1 - 2y^2 + y^4$ |
| 3 | $\frac{3}{16}\sqrt{5\cdot 7\cdot 11} = 3.6790157$ | $y - 2y^3 + y^5$ |
| 4 | $\frac{3}{32}\sqrt{7\cdot 13} = 0.8943180$ | $[-1 + 11y^2]\chi_2$ |
| 5 | $\frac{9}{32}\sqrt{5\cdot 7\cdot 11} = 5.5185234$ | $[-1 + \frac{13}{3}y^2]\chi_3$ |
| 6 | $\frac{3}{128}\sqrt{7\cdot 11\cdot 17} = 0.8479712$ | $[1 - 26y^2 + 65y^4]\chi_2$ |
| 7 | $\frac{3}{128}\sqrt{5\cdot 7\cdot 11\cdot 13\cdot 19} = 7.2275337$ | $[1 - 10y^2 + 17y^4]\chi_3$ |
| 8 | $\frac{3}{256}\sqrt{5\cdot 7\cdot 11\cdot 13} = 0.8290550$ | $[-1 + 45y^2 - 255y^4 + 323y^6]\chi_2$ |
| 9 | $\frac{3\cdot 5}{256}\sqrt{7\cdot 11\cdot 13\cdot 23} = 8.8906241$ | $[-1 + 17y^2 - \frac{323}{5}y^4 + \frac{323}{5}y^6]\chi_3$ |

## TABLE IB

*Expansion coefficients $\langle y^k f_n \rangle$*

| $k$ | $n = 2$ | 4 | 6 | 8 |
|---|---|---|---|---|
| 0 | 1.1832160 | .5451081 | .3445724 | .2450079 |
| 2 | .1690309 | .3634054 | .2819229 | .2167377 |
| 4 | .0563436 | .1817027 | .2192734 | .1884676 |
| 6 | .0256107 | .1016519 | .1566238 | .1601975′ |
| 8 | .0137904 | .0622618 | .1121171 | .1319273 |
| 10 | .0082742 | .0408102 | .0819338 | .1073915 |
| $k$ | $n = 3$ | 5 | 7 | 9 |
| 1 | .5606119 | .3737413 | .2669911 | .2021090 |
| 3 | .1868706 | .2718118 | .2259156 | .1818981 |
| 5 | .0849412 | .1698824 | .1848400 | .1616872 |
| 7 | .0457376 | .1097702 | .1437645′ | .1414763 |
| 9 | .0274425 | .0742563 | .1108277 | .1212654 |
| 11 | .0177569 | .0523362 | .0860298 | .1027956 |

## TABLE II

*Approximate eigenvalues, $\alpha_n + i\beta_n$*

| $n$ | $\alpha_n$ | $\beta_n$ |
|---|---|---|
| 2 | 2.075149 | 1.142910 |
| 3 | 3.655963 | 1.538202 |
| 4 | 5.258543 | 2.062104 |
| 5 | 6.942082 | 2.681885 |
| 6 | 8.728814 | 3.404732 |
| 7 | 10.627227 | 4.235321 |
| 8 | 12.640404 | 5.174922 |
| 9 | 14.769022 | 6.222861 |

TABLE III

| $y$ | $n=2$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| | | | | *Self-equilibrating polynomials, $f_n(y)$* | | | | |
| 0 | 1.10927 | 0.00000 | −0.89432 | 0.00000 | 0.84797 | 0.00000 | −0.82905 | 0.00000 |
| 0.05 | 1.10373 | 0.18303 | −0.86538 | −0.27157 | 0.78924 | 0.35062 | −0.73342 | −0.42369 |
| 0.10 | 1.08719 | 0.36058 | −0.78010 | −0.51743 | 0.62041 | 0.63874 | −0.46736 | −0.72881 |
| 0.15 | 1.05991 | 0.52730 | −0.64303 | −0.71383 | 0.36291 | 0.81173 | −0.08945′ | −0.82759 |
| 0.20 | 1.02230 | 0.67812 | −0.46155 | −0.84086 | 0.05002 | 0.83554 | 0.31530 | −0.68699 |
| 0.25 | 0.97494 | 0.80838 | −0.24563 | −0.88416 | −0.27657 | 0.70099 | 0.65234 | −0.34005 |
| 0.30 | 0.91858 | 0.91398 | −0.00741 | −0.83629 | −0.57124 | 0.42680 | 0.83756 | 0.11890 |
| 0.35 | 0.85414 | 0.99150 | 0.23930 | −0.69777 | −0.78980 | 0.05864 | 0.81691 | 0.55552 |
| 0.40 | 0.78270 | 1.03837 | 0.47958 | −0.47765′ | −0.89510 | −0.33618 | 0.58206 | 0.83018 |
| 0.45 | 0.70550′ | 1.05294 | 0.69819 | −0.19348 | −0.86268 | −0.67826 | 0.17823 | 0.83949 |
| 0.50 | 0.62396 | 1.03472 | 0.88034 | 0.12934 | −0.68566 | −0.88933 | −0.29875 | 0.55480 |
| 0.55 | 0.53966 | 0.98443 | 1.01267 | 0.45899 | −0.37834 | −0.90778 | −0.71825′ | 0.04607 |
| 0.60 | 0.45435 | 0.90415 | 1.08429 | 0.75949 | 0.02223 | −0.70481 | −0.94339 | −0.52042 |
| 0.65 | 0.36995′ | 0.79753 | 1.08791 | 0.99393 | 0.45755′ | −0.29830 | −0.86993 | −0.91921 |
| 0.70 | 0.28852 | 0.66984 | 1.02117 | 1.12868 | 0.83278 | 0.23910 | −0.46898 | −0.93940 |
| 0.75 | 0.21232 | 0.52814 | 0.88799 | 1.13880 | 1.12664 | 0.78221 | 0.17709 | −0.48490 |
| 0.80 | 0.14376 | 0.38144 | 0.70006 | 1.01463 | 1.20711 | 1.17139 | 0.86220 | 0.32663 |
| 0.85 | 0.08542 | 0.24081 | 0.47846 | 0.76969 | 1.05428 | 1.25324 | 1.29091 | 1.12009 |
| 0.90 | 0.04004 | 0.11953 | 0.25537 | 0.45004 | 0.69141 | 0.95190 | 1.19117 | 1.36254 |
| 0.95 | 0.01054 | 0.03322 | 0.07590 | 0.14507 | 0.24568 | 0.37998 | 0.54655′ | 0.73966 |
| 0.98 | 0.00174 | 0.00565 | 0.01341 | 0.02681 | 0.04785′ | 0.07860 | 0.12109 | 0.17717 |
| 1.00 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| | | | | *First derivatives, $f_n'(y)$* | | | | |
| 0 | 0.00000 | 3.67902 | 0.00000 | −5.51852 | 0.00000 | 7.22753 | 0.0000 | −8.8906 |
| 0.05 | −0.22130 | 3.62395′ | 1.15235′ | −5.25806 | −2.32454 | 6.58560 | 3.7545′ | −7.6512 |
| 0.10 | −0.43927 | 3.46011 | 2.24354 | −4.49651 | −4.35629 | 4.76073 | 6.6888 | −4.2528 |
| 0.15 | −0.65058 | 3.19166 | 3.21464 | −3.29248 | −5.83166 | 2.04336 | 8.1387 | 0.4157 |
| 0.20 | −0.85192 | 2.82548 | 4.01112 | −1.74050′ | −6.54368 | −1.12258 | 7.7294 | 5.0965 |

TABLE III—*First derivatives, $f'_n(y)$—Continued*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.25 | −1.03994 | 2.37124 | 4.58513 | 0.03368 | −6.36599 | −4.19684 | 5.4651 | 8.4595′ |
| 0.30 | −1.21132 | 1.84135′ | 4.89763 | 1.87968 | −5.27163 | −6.61676 | 1.7579 | 9.4389 |
| 0.35 | −1.36273 | 1.25098 | 4.92067 | 3.63096 | −3.34481 | −7.88195′ | −2.6144 | 7.5460 |
| 0.40 | −1.49085 | 0.61807 | 4.63958 | 5.11519 | −0.78410 | −7.64166 | −6.6279 | 3.0823 |
| 0.45 | −1.59235′ | −0.03668 | 4.05518 | 6.16647 | 2.10488 | −5.77553 | −9.2005 | −2.8139 |
| 0.50 | −1.66390 | −0.68982 | 3.18601 | 6.63947 | 4.92883 | −2.45623 | −9.4337 | −8.3402 |
| 0.55 | −1.70217 | −1.31513 | 2.07050 | 6.42546 | 7.24019 | 1.81902 | −6.8767 | −11.4424 |
| 0.60 | −1.70383 | −1.88366 | 0.76926 | 5.47014 | 8.58597 | 6.24289 | −1.7669 | −10.4406 |
| 0.65 | −1.66556 | −2.36365 | −0.63280 | 3.79346 | 8.57416 | 9.77477 | 4.8143 | −4.7734 |
| 0.70 | −1.58403 | −2.72063 | −2.02418 | 1.51126 | 6.95906 | 11.29847 | 10.9544 | 4.3217 |
| 0.75 | −1.45591 | −2.91734 | −3.26461 | −1.14116 | 3.74755 | 9.88062 | 14.1916 | 13.4846 |
| 0.80 | −1.27787 | −2.91378 | −4.18283 | −3,78368 | −0.67208 | 5.15170 | 12.1924 | 17.6993 |
| 0.85 | −1.04659 | −2.66717 | −4.57439 | −5.86398 | −5.37706 | −2.16691 | 3.9742 | 12.1463 |
| 0.90 | −0.75874 | −2.13199 | −4.19941 | −6.62843 | −8.73093 | −9.56452 | −8.1501 | −3.7649 |
| 0.95 | −0.41098 | −1.25995′ | −2.78040 | −5.09092 | −8.17662 | −11.84440 | −15.7012 | −19.1634 |
| 0.98 | −0.17219 | −0.55391 | −1.29756 | −2.55494 | −4.47895 | −7.20701 | −10.8428 | −15.4382 |
| 1.0 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.0000 |
| | | | *Second derivatives, $f''_n(y)$* | | | | | |
| 0 | −4.43706 | 0.0000 | 23.2523 | 0.000 | −47.486 | 0.000 | 77.931 | 0.000 |
| 0.05 | −4.40378 | −2.1982 | 22.6370 | 10.352 | −44.509 | −25.337 | 69.461 | 48.487 |
| 0.10 | −4.30395′ | −4.3412 | 20.8135′ | 19.914 | −35.873 | −46.678 | 45.651 | 84.418 |
| 0.15 | −4.13756 | −6.3739 | 17.8480 | 27.931 | −22.444 | −60.518 | 11.066 | 98.026 |
| 0.20 | −3.90461 | −8.2410 | 13.8512 | 33.727 | −5.609 | −64.294 | −27.434 | 84.658 |
| 0.25 | −3.60511 | −9.8874 | 8.9781 | 36.736 | 12.811 | −56.772 | −61.751 | 46.200 |
| 0.30 | −3.23905 | −11.2578 | 3.4279 | 36.545 | 30.684 | −38.327 | −83.901 | −8.726 |
| 0.35 | −2.80644 | −12.2971 | −2.5559 | 32.928 | 45.724 | −11.084 | −87.516 | −66.002 |
| 0.40 | −2.30727 | −12.9501 | −8.6856 | 25.884 | 55.679 | 21.118 | −69.336 | −108.696 |
| 0.45 | −1.74155′ | −13.1617 | −14.6292 | 15.678 | 58.558 | 52.963 | −30.483 | −121.006 |
| 0.50 | −1.10927 | −12.8766 | −20.0104 | 2.874 | 52.892 | 78.148 | 22.741 | −92.983 |
| 0.55 | −0.41043 | −12.0396 | −24.4086 | −11.622 | 38.030 | 90.147 | 78.806 | −25.249 |
| 0.60 | 0.35496 | −10.5956 | −27.3590 | −26.531 | 14.466 | 83.288 | 121.982 | 67.431 |
| 0.65 | 1.18691 | −8.4893 | −28.3525 | −40.158 | −15.793 | 54.182 | 134.804 | 155.261 |
| 0.70 | 2.08542 | −5.6657 | −26.8358 | −50.354 | −48.831 | 3.529 | 102.375′ | 197.215′ |
| 0.75 | 3.05048 | −2.0694 | −22.2112 | −54.485′ | −78.386 | −61.649 | 18.935′ | 152.001 |

TABLE III—*Second derivatives, $f_n''(y)$—Continued*

| $y$ | $n=2$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| 0.80 | 4.08210 | 2.3546 | −13.8369 | −49.385′ | −95.383 | −125.334 | −102.817 | −0.775 |
| 0.85 | 5.18027 | 7.6616 | −1.0266 | −31.326 | −87.430 | −159.339 | −219.146 | −225.663 |
| 0.90 | 6.34500′ | 13.9067 | 16.9500 | 4.024 | −38.280 | −119.617 | −238.893 | −378.379 |
| 0.95 | 7.57628 | 21.1451 | 40.8677 | 61.636 | 72.743 | 57.984 | −2.694 | −129.902 |
| 0.98 | 8.34700 | 25.9880 | 58.4089 | 109.210 | 179.524 | 266.594 | 362.475′ | 453.012 |
| 1.00 | 8.87412 | 29.4321 | 71.5454 | 147.161 | 271.351 | 462.562 | 742.833 | 1138.000 |

*Integrals* $f_n(y) = \int_0^y f_n(y)\,dy$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| .05 | .055371 | .004587 | −.044233 | −.006844 | .041415 | .008900 | −.039847 | −.010852 |
| .10 | .110189 | .018212 | −.085597 | −.026728 | .077081 | .034015′ | −.070480 | −.040375 |
| .15 | .163911 | .040465′ | −.121378 | −.057760 | .101972 | .070844 | −.084704 | −.080262 |
| .20 | .216008 | .070676 | −.149159 | −.096952 | .112444 | .112688 | −.078972 | −.119107 |
| .25 | .265978 | .107933 | −.166958 | −.140447 | .106743 | .151743 | −.054308 | −.145487 |
| .30 | .313352 | .151103 | −.173350′ | −.183844 | .085320 | .180444 | −.016285 | −.151222 |
| .35 | .357701 | .198863 | −.167557 | −.222561 | .050892 | .192845′ | .025991 | −.133966 |
| .40 | .398649 | .249741 | −.149526 | −.252256 | .008235′ | .185857 | .061805 | −.098390 |
| .45 | .435875 | .302161 | −.119960 | −.269254 | −.036313 | .160107 | .081352 | −.055413 |
| .50 | .469127 | .354489 | −.080316 | −.270956 | −.075612 | .120224 | .078389 | −.019398 |
| .55 | .498225 | .405098 | −.032758 | −.256204 | −.102695 | .074403 | .052431 | −.003724 |
| .60 | .523076 | .452431 | .019938 | −.225542 | −.111880 | .033163 | .009821 | −.015789 |
| .65 | .543676 | .495073 | .074535 | −.181357 | −.099884 | .007345′ | −.036890 | −.052964 |
| .70 | .560120 | .531832 | .127552 | −.127816 | −.066789 | .005544 | −.071651 | −.101336 |
| .75 | .572615′ | .561822 | .175540 | −.070575 | −.016633 | .031371 | −.079631 | −.138871 |
| .80 | .581480 | .584561 | .215433 | −.016187 | .042635′ | .081198 | −.053236 | −.143724 |
| .85 | .587161 | .600066 | .244979 | .028857 | .100155 | .143347 | .002310 | −.106404 |
| .90 | .590238 | .608964 | .263247 | .059513 | .144505 | .200034 | .066916 | −.040990 |
| .95 | .591430 | .612601 | .271235′ | .074075′ | .167830 | .233837 | .111995′ | .014885 |
| .98 | .591596 | .613131 | .272463 | .076463 | .171957 | .240371 | .121653 | .028375 |
| 1.00 | .591608 | .613169 | .272554 | .076646 | .172286 | .240918 | .122504 | .029635 |

and satisfy the boundary conditions

$$f_n(\pm 1) = 0, \qquad f_n'(\pm 1) = 0 \tag{5}$$

The expansion coefficients of $y^k$, $k = 0$ to 10, with respect to $f_n(y)$ are listed in Table Ib. (The coefficients are zero for $k + n =$ odd number.)

Then, on denoting

$$G_k(x, y) = f_k(y)g_k(x), \qquad H_k(x, y) = f_k(y)h_k(x)$$

$$g_k(x) = e^{-\alpha_k x}\left(\cos \beta_k x + \frac{\alpha_k}{\beta_k} \sin \beta_k x,\right) \qquad h_k(x) = e^{-\alpha_k x} \frac{\sin \beta_k x}{\beta_k} \tag{6a}$$

the approximate stress functions may be written as

$$\text{Case (a)} \qquad \varphi(x, y) = \sum \langle \bar{\varphi}, f_k \rangle G_k$$

(6b)

$$\text{Case (b)} \qquad \varphi(x, y) = \sum \langle \bar{\varphi}_{,x}, f_k \rangle H_k$$

From these the stresses are obtained, in the usual manner, by differentiation, Case (b)

$$\begin{aligned} \sigma_x(x, y) = \varphi_{,yy} &= \sum \langle \bar{\varphi}_{,x}, f_k \rangle f_k'' h_k \\ \sigma_y(x, y) = \varphi_{,xx} &= \sum \langle \bar{\varphi}_{,x}, f_k \rangle f_k h_k'' \\ \tau(x, y) = -\varphi_{,xy} &= -\sum \langle \bar{\varphi}_{,x}, f_k \rangle f_k' h_k' \end{aligned} \tag{7}$$

Case (a) = similar

In the above formulas (6) the approximate eigenvalues

$$\alpha_n + i\beta_n = [\langle f_n'^2 \rangle + i\sqrt{\langle f_n''^2 \rangle - \langle f_n'^2 \rangle^2}\,]^{\frac{1}{2}} \tag{8}$$

have values as tabulated in Table II for $n = 2$ to 9; and the functions $f_n$, $f_n'$, $f_n''$, $\bar{f}_n$; $g_n$, $g_n'$, $g_n''$, $\bar{g}_n$; $h_n$, $h_n'$, $h_n''$, $\bar{h}_n$ have values as tabulated in Tables III and IV. We denote as usual

$$f_n' = df_n/dy, \qquad \bar{f}_n = \int_0^y f_n\, dy, \qquad g_n' = dg_n/dx, \qquad \bar{g}_n = \int_\infty^x g_n\, dx \tag{9}$$

Enough decimal places are retained in the entries so that the largest term in any one column appears to six digit accuracy. Prime after a 5 (or a 50) indicates that the digit was arrived at by rounding *up*. Table IV, case $n = 2$, includes also a column $g^{IV}$.

The tables will permit rapid solution of a variety of endproblems. While the functions $G_k$, $H_k$ are approximate, comparison with available (very laborious)

TABLE IV

| $x$ | $z^{IV}$ | $z$ | $z'$ | $z''$ | $\bar{z}$ | $h$ | $h'$ | $h''$ | $\bar{h}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | *Functions $g_n$, $h_n$ and their derivatives $n = 2$* | | | | | |
| 0 | −65.1749 | 1.000000 | 0.000000 | −5.61249 | −0.739476 | 0.000000 | 1.000000 | −4.15030 | −0.178174 |
| 0.1 | −52.2497 | 0.975560 | −0.455079 | −3.58660 | −0.640319 | 0.081083 | 0.639040 | −3.10728 | −0.173820 |
| 0.2 | −41.3261 | 0.914817 | −0.734770 | −2.08488 | −0.545568 | 0.130917 | 0.371472 | −2.27649 | −0.162997 |
| 0.4 | −24.7395′ | 0.740697 | −0.945127 | −0.23459 | −0.379330 | 0.168397 | 0.041798 | −1.11860 | −0.131973 |
| 0.6 | −13.8030 | 0.553869 | −0.895330 | 0.60731 | −0.250048 | 0.159525′ | −0.108207 | −0.44624 | −0.098685 |
| 0.8 | −6.9677 | 0.389474 | −0.739557 | 0.88346 | −0.156237 | 0.131770 | −0.157410 | −0.08626 | −0.069394 |
| 1.0 | −2.9440 | 0.259477 | −0.560899 | 0.87158 | −0.091939 | 0.099938 | −0.155294 | 0.08362 | −0.046232 |
| 1.2 | −0.7488 | 0.163943 | −0.399015′ | 0.73590 | −0.050138 | 0.071094 | −0.131119 | 0.14517 | −0.029210 |
| 1.4 | 0.3205′ | 0.097740 | −0.268683 | 0.56655 | −0.024404 | 0.047872 | −0.100945′ | 0.15027 | −0.017415′ |
| 1.6 | 0.7387 | 0.054239 | −0.171625′ | 0.40788 | −0.009529 | 0.030579 | −0.072673 | 0.12999 | −0.009664 |
| 1.8 | 0.8105′ | 0.027150 | −0.103608 | 0.27762 | −0.001617 | 0.018460 | −0.049465′ | 0.10169 | −0.004837 |
| 2.0 | 0.7204 | 0.011274 | −0.058436 | 0.17926 | 0.002075 | 0.010412 | −0.031939 | 0.07412 | −0.002009 |
| 2.4 | 0.4233 | −0.001490 | −0.013097 | 0.06272 | 0.003436 | 0.002334 | −0.011175 | 0.03328 | 0.000265 |
| 2.8 | 0.1943 | −0.003309 | 0.000861 | 0.01500 | 0.002294 | −0.000153 | −0.002673 | 0.01195 | 0.000590 |
| 3.2 | 0.0715 | −0.002306 | 0.003164 | −0.00019 | 0.001142 | −0.000564 | 0.000033 | 0.00303 | 0.000411 |
| 3.6 | 0.0190 | −0.001175 | 0.002312 | −0.00300 | 0.000457 | −0.000412 | 0.000534 | 0.00009 | 0.000209 |
| 4.0 | 0.0013 | −0.000481 | 0.001208 | −0.00231 | 0.000141 | −0.000215 | 0.000412 | −0.00050 | 0.000086 |

| $x$ | $z$ | $z'$ | $z''$ | $\bar{z}$ | $h$ | $h'$ | $h''$ | $\bar{h}$ |
|---|---|---|---|---|---|---|---|---|
| | | | | *$n = 3$* | | | | |
| 0 | 1.00000 | 0.00000 | −15.73213 | −0.464777 | 0.0000000 | 1.000000 | −7.31193 | −0.0635642 |
| 0.1 | 0.938238 | −1.08717 | −6.81116 | −0.366966 | 0.0691050 | 0.432946 | −4.25284 | −0.0596383 |
| 0.2 | 0.805159 | −1.49071 | −1.76692 | −0.279463 | 0.0947555 | 0.112313 | −2.31193 | −0.0511793 |
| 0.4 | 0.507029 | −1.36768 | 2.02377 | −0.148719 | 0.0869357 | −0.128639 | −0.42709 | −0.0322289 |
| 0.6 | 0.278643 | −0.90944 | 2.26610 | −0.071699 | 0.0578078 | −0.144043 | 0.14379 | −0.0177117 |
| 0.8 | 0.138178 | −0.51752 | 1.61022 | −0.031326 | 0.0328956 | −0.102352 | 0.23087 | −0.0087832 |
| 1.0 | 0.062217 | −0.26411 | 0.95231 | −0.012129 | 0.0167877 | −0.060533 | 0.17851 | −0.0039548 |
| 1.2 | 0.025069 | −0.12241 | 0.50068 | −0.003870 | 0.0077809 | −0.031825′ | 0.11029 | −0.0015935′ |

TABLE IV—$n = 3$—*Continued*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1.4 | 0.008585 | −0.05112 | 0.23870 | −0.000741 | 0.0032493 | −0.015173 | 0.05983 | −0.0005457 |
| 1.6 | 0.002069 | −0.01854 | 0.10300 | 0.000217 | 0.0011785′ | −0.006547 | 0.02933 | −0.0001315 |
| 1.8 | −0.000091 | −0.00517 | 0.03922 | 0.000371 | 0.0003284 | −0.002493 | 0.01306 | 0.0000058 |
| 2.0 | −0.000563 | −0.00045′ | 0.01210 | 0.000290 | 0.0000283 | −0.000769 | 0.00518 | 0.0000358 |
| 2.4 | −0.000324 | 0.00083 | −0.00094 | 0.000098 | −0.0000526 | +0.000060 | 0.00039 | 0.0000206 |
| 2.8 | −0.000092 | 0.00034 | −0.00101 | 0.000022 | −0.0000214 | +0.000064 | −0.00013 | 0.0000058 |
| 3.2 | −0.000018 | 0.00008 | −0.00033 | 0.000003 | −0.0000053 | +0.000021 | −0.00007 | 0.0000011 |
| 3.6 | −0.000002 | 0.00001 | −0.00007 | −0.000000 | −0.0000008 | +0.000004 | −0.00002 | 0.0000001 |
| | | | | $n = 4$ | | | | |
| 0 | 1.000000 | 0.00000 | −31.9045 | −0.329642 | 0.0000000 | 1.000000 | −10.51709 | −0.0313435′ |
| 0.1 | 0.887140 | −1.87239 | −8.6117 | −0.233752 | 0.0586873 | 0.269921 | −4.71117 | −0.0278061 |
| 0.2 | 0.677126 | −2.16645 | 1.1814 | −0.155305 | 0.0679042 | −0.037028 | −1.77702 | −0.0212235′ |
| 0.4 | 0.311388 | −1.38673 | 4.6498 | −0.059181 | 0.0434651 | −0.145740 | 0.14602 | −0.0097600 |
| 0.6 | 0.116684 | −0.62326 | 2.8321 | −0.018929 | 0.0195351 | −0.088769 | 0.31033 | −0.0036573 |
| 0.8 | 0.036688 | −0.22971 | 1.2454 | −0.004894 | 0.0072000 | −0.039035 | 0.18082 | −0.0011499 |
| 1.0 | 0.009244 | −0.07097 | 0.4515 | −0.000822 | 0.0022246 | −0.014153 | 0.07787 | −0.0002897 |
| 1.2 | 0.001440 | −0.01740 | 0.1370 | 0.000071 | 0.0005453 | −0.004296 | 0.02778 | −0.0000451 |
| 1.4 | −0.000207 | −0.00247 | 0.0326 | 0.000146 | 0.0000775 | −0.001022 | 0.00828 | 0.0000065′ |
| 1.6 | −0.000308 | 0.00054 | 0.0042 | 0.000085′ | −0.0000169 | −0.000130 | 0.00191 | 0.0000096 |
| 1.8 | −0.000172 | 0.00065′ | −0.0013 | 0.000036 | −0.0000203 | 0.000041 | 0.00021 | 0.0000054 |
| 2.0 | −0.000072 | 0.00035′ | −0.0014 | 0.000013 | −0.0000109 | 0.000042 | −0.00010 | 0.0000023 |
| 2.4 | −0.000007 | 0.00005 | −0.0003 | 0.000001 | −0.0000016 | 0.000009 | −0.00004 | 0.0000002 |
| 2.8 | 0.000000 | 0.00000 | −0.0000 | 0.000000 | −0.0000001 | 0.000001 | −0.00001 | 0.0000000 |
| | | | | $n = 5$ | | | | |
| 0 | 1.000000 | 0.00000 | −55.3850 | −0.250685′ | 0.0000000 | 1.000000 | −13.88416 | −0.0180554 |
| 0.05 | 0.944965′ | −1.95125 | −25.2454 | −0.201657 | 0.0352307 | 0.455816 | −8.27987 | −0.0170617 |
| 0.10 | 0.824208 | −2.73327 | −7.6096 | −0.157266 | 0.0493504 | 0.139019 | −4.66344 | −0.0148814 |
| 0.15 | 0.682553 | −2.85411 | 1.8237 | −0.119573 | 0.0515322 | −0.032928 | −2.93693 | −0.0123238 |
| 0.20 | 0.544431 | −2.63275 | 6.4003 | −0.088945′ | 0.0475355 | −0.115560 | −1.02830 | −0.0098299 |
| 0.30 | 0.318800 | −1.85410 | 8.0859 | −0.046442 | 0.0334766 | −0.145994 | 0.17291 | −0.0057561 |
| 0.40 | 0.171257 | −1.12912 | 6.1919 | −0.022545′ | 0.0203868 | −0.111797 | 0.42308 | −0.0030921 |

TABLE IV—$n = 5$—*Continued*

| $z$ | $\xi$ | $\xi'$ | $\xi''$ | $\bar{\xi}$ | $h$ | $h'$ | $h''$ | $\bar{h}$ |
|---|---|---|---|---|---|---|---|---|
| 0.60 | 0.039564 | −0.32040 | 2.2572 | −0.004133 | 0.0057849 | −0.040754 | 0.24544 | −0.0007143 |
| 0.80 | 0.006310 | −0.06714 | 0.5827 | −0.000370 | 0.0012122 | −0.010521 | 0.07893 | −0.0001139 |
| 1.0 | 0.000244 | −0.00886 | 0.1094 | 0.000099 | 0.0001599 | −0.001976 | 0.01858 | −0.0000044 |
| 1.2 | −0.000288 | 0.00038 | 0.0107 | 0.000065 | −0.0000069 | −0.000193 | 0.00305 | 0.0000052 |
| 1.4 | −0.000139 | 0.00071 | −0.0022 | 0.000022 | −0.0000129 | 0.000040 | 0.00015 | 0.0000025 |
| 1.6 | −0.000042 | 0.00028 | −0.0016 | 0.000005 | −0.0000051 | 0.000029 | −0.00012 | 0.0000008 |
| 1.8 | −0.000009 | 0.00008 | −0.0006 | 0.000001 | −0.0000014 | 0.000010 | −0.00006 | 0.0000002 |
| 2.0 | −0.000001 | 0.00002 | −0.0001 | 0.000000 | −0.0000003 | 0.000002 | −0.00002 | 0.0000000 |
| | | | | $n = 6$ | | | | |
| 0 | 1.000000 | 0.00000 | −87.7844 | −0.198869 | 0.0000000 | 1.000000 | −17.45763 | −0.0113915 |
| 0.05 | 0.917715′ | −2.82321 | −31.2746 | −0.150345′ | 0.0321607 | 0.356266 | −9.04276 | −0.0104541 |
| 0.10 | 0.751406 | −3.59673 | −3.1714 | −0.108459 | 0.0409723 | 0.036127 | −4.22742 | −0.0085596 |
| 0.15 | 0.573903 | −3.40277 | 9.0244 | −0.075369 | 0.0387628 | −0.102802 | −1.60808 | −0.0065376 |
| 0.20 | 0.417245′ | −2.83254 | 12.8219 | −0.050710 | 0.0322670 | −0.146061 | −0.28266 | −0.0047530 |
| 0.30 | 0.197465′ | −1.60305′ | 10.6511 | −0.021009 | 0.0182612 | −0.121332 | 0.51513 | −0.0022494 |
| 0.40 | 0.082695 | −0.76813 | 6.1505′ | −0.007695 | 0.0087502 | −0.070063 | 0.45501 | −0.0009420 |
| 0.60 | 0.009719 | −0.12204 | 1.2774 | −0.000542 | 0.0013902 | −0.014552 | 0.13200 | −0.0001107 |
| 0.80 | 0.000117 | −0.00970 | 0.1591 | 0.000087 | 0.0001105 | −0.001812 | 0.02194 | −0.0000013 |
| 1.0 | −0.000264 | 0.00109 | 0.0042 | 0.000040 | −0.0000124 | −0.000048 | 0.00193 | 0.0000030 |
| 1.2 | −0.000075′ | 0.00059 | −0.0037 | 0.000008 | −0.0000067 | 0.000042 | −0.00014 | 0.0000009 |
| 1.4 | −0.000012 | 0.00012 | −0.0011 | 0.000001 | −0.0000014 | 0.000013 | −0.00010 | 0.0000001 |
| 1.6 | −0.000001 | 0.00002 | −0.0002 | 0.000000 | 0.0000002 | 0.000002 | −0.00002 | 0.0000000 |
| | | | | $n = 7$ | | | | |
| 0 | 1.000000 | 0.00000 | −130.8759 | −0.162402 | 0.0000000 | 1.000000 | −21.2545′ | −0.0076408 |
| 0.05 | 0.884682 | −3.81779 | −34.6385′ | −0.114503 | 0.0291711 | 0.264666 | −9.4431 | −0.0067597 |
| 0.10 | 0.671293 | −4.38797 | 5.4079 | −0.075491 | 0.0335277 | −0.041321 | −3.5097 | −0.0051292 |
| 0.15 | 0.465879 | −3.72421 | 18.1836 | −0.047203 | 0.0284560 | −0.138938 | −0.7712 | −0.0035597 |

TABLE IV—$n = 7$—*Continued*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.20 | 0.303515 | −2.76430 | 19.0308 | −0.028170 | 0.0211215 | −0.145411 | 0.3263 | −0.0023191 |
| 0.30 | 0.111066 | −1.21758 | 11.3432 | −0.008734 | 0.0093033 | −0.086671 | 0.6246 | −0.0008486 |
| 0.40 | 0.033735′ | −0.43703 | 4.8740 | −0.002139 | 0.0033393 | −0.037241 | 0.3545 | −0.0002578 |
| 0.60 | 0.001008 | −0.02970 | 0.4994 | 0.000063 | 0.0002269 | −0.003816 | 0.0514 | −0.0000077 |
| 0.80 | −0.000321 | 0.00153 | 0.0095 | 0.000040 | −0.0000117 | −0.000073 | 0.0031 | 0.0000024 |
| 1.0 | −0.000065′ | 0.00067 | −0.0056 | 0.000005 | −0.0000051 | 0.000043 | −0.0002 | 0.0000004 |
| 1.2 | −0.000006 | 0.00008 | −0.0010 | 0.000000 | −0.0000006 | 0.000008 | −0.0001 | 0.0000000 |
| | | | | $n = 8$ | | | | |
| 0 | 1.000000 | 0.00000 | −186.5596 | −0.1355106 | 0.0000000 | 1.000000 | −25.2808 | −0.0053602 |
| 0.05 | 0.846018 | −4.90286 | −33.8846 | −0.0883641 | 0.0262804 | 0.181629 | −9.4946 | −0.0045348 |
| 0.10 | 0.586897 | −5.03840 | 17.8833 | −0.0525239 | 0.0270069 | −0.095859 | −2.6150 | −0.0031459 |
| 0.15 | 0.364115′ | −3.79259 | 27.9506 | −0.0290123 | 0.0203291 | −0.149821 | −0.0050′ | −0.0019517 |
| 0.20 | 0.208377 | −2.47404 | 23.6711 | −0.0149759 | 0.0132614 | −0.126882 | 0.7336 | −0.0011169 |
| 0.30 | 0.055479 | −0.81273 | 10.1961 | −0.0031617 | 0.0043564 | −0.054653 | 0.5690 | −0.0002974 |
| 0.40 | 0.010611 | −0.20161 | 3.1174 | −0.0003572 | 0.0010807 | −0.016710 | 0.2208 | −0.0000569 |
| 0.60 | −0.000463 | −0.00067 | 0.1033 | 0.0000663 | 0.0000036 | −0.000554 | 0.0133 | 0.0000025′ |
| 0.80 | −0.000105 | 0.00123 | −0.0114 | 0.0000077 | −0.0000066 | 0.000061 | −0.0003 | 0.0000006 |
| 1.0 | −0.000006 | 0.00011 | −0.0016 | 0.0000002 | −0.0000006 | 0.000008 | −0.0001 | 0.0000000 |
| | | | | $n = 9$ | | | | |
| 0 | 1.000000 | 0.00000 | −256.848 | −0.1150020 | 0.0000000 | 1.000000 | −29.5380 | −0.0038934 |
| 0.05 | 0.802115 | −6.03827 | −27.663 | −0.0687358 | 0.0235091 | 0.107703 | −9.2196 | −0.0031230 |
| 0.10 | 0.501435 | −5.49372 | 33.481 | −0.0362771 | 0.0213890 | −0.130354 | −1.6433 | −0.0019523 |
| 0.15 | 0.273057 | −3.61948 | 36.779 | −0.0173101 | 0.0140919 | −0.143192 | 0.6101 | −0.0010631 |
| 0.20 | 0.133932 | −2.03863 | 25.817 | −0.0074654 | 0.0079371 | −0.100513 | 0.9303 | −0.0005215′ |
| 0.30 | 0.023554 | −0.47003 | 7.834 | −0.0008787 | 0.0018300 | −0.030501 | 0.4309 | −0.0000917 |
| 0.40 | 0.001757 | −0.06812 | 1.561 | 0.0000632 | 0.0002652 | −0.006078 | 0.1114 | −0.0000068 |
| 0.60 | −0.000305 | 0.00326 | −0.018 | 0.0000224 | −0.0000127 | 0.000070 | 0.0012 | 0.0000012 |
| 0.80 | −0.000015 | 0.00028 | −0.005′ | 0.0000006 | −0.0000011 | 0.000019 | −0.0003 | 0.0000001 |

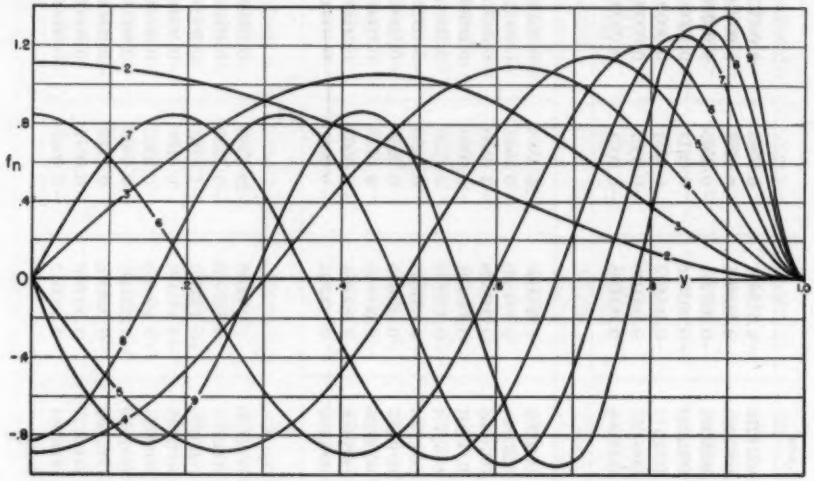



FIG. 2

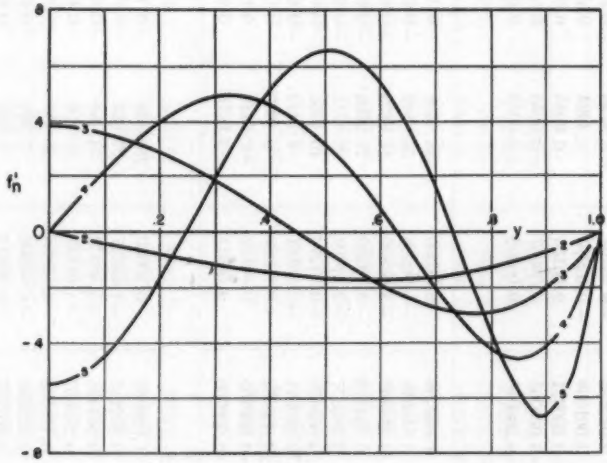



FIG. 3a

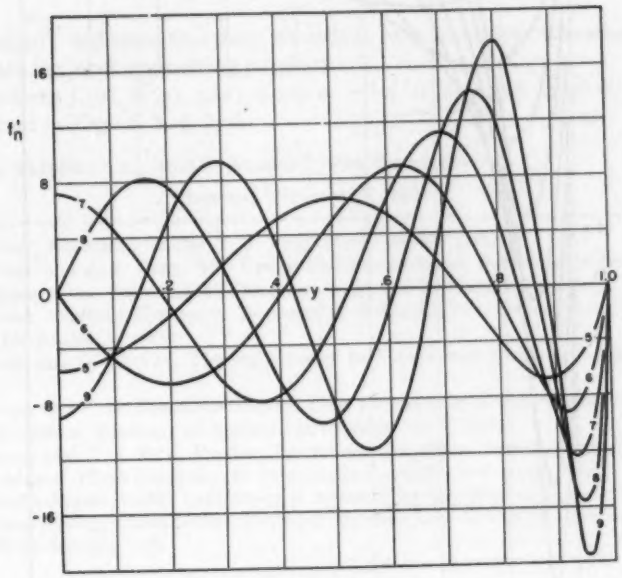

16
8
0
-8
-16
$f_n'$
.2
.4
y
.6
.8
1.0
7
8
6
5
9
5
6
7
8
9

FIG. 3b

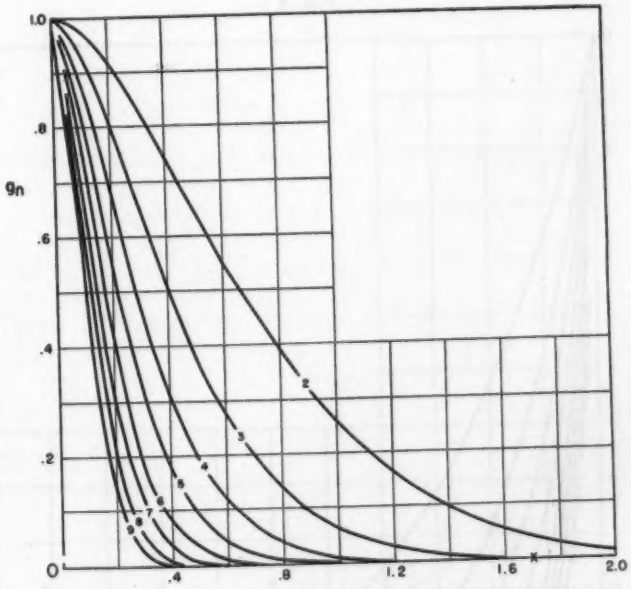

1.0
.8
.6
.4
.2
0
$g_n$
2
3
4
5
6
7
8
9
.4
.8
1.2
1.6
2.0
x

FIG. 4

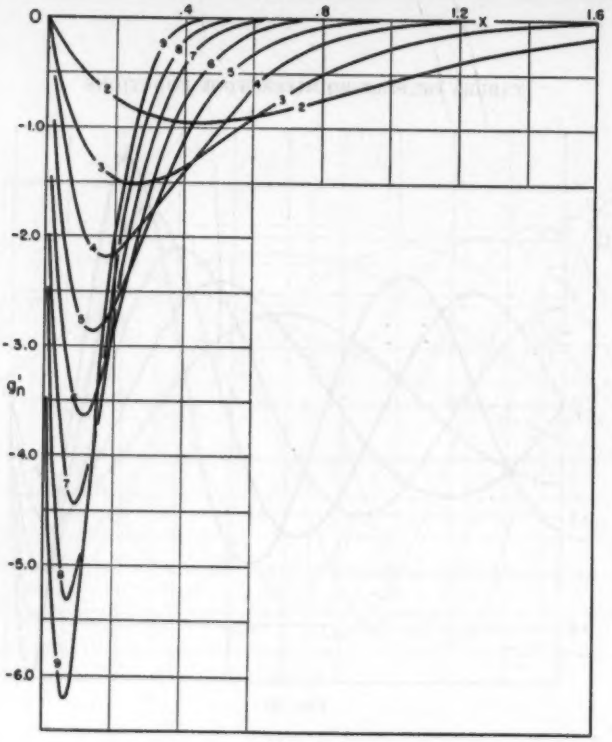

0
.4
.8
1.2
X
1.6
-1.0
-2.0
-3.0
$g_n'$
-4.0
-5.0
-6.0
2
3
4
5
6
7
8
9

FIG. 5

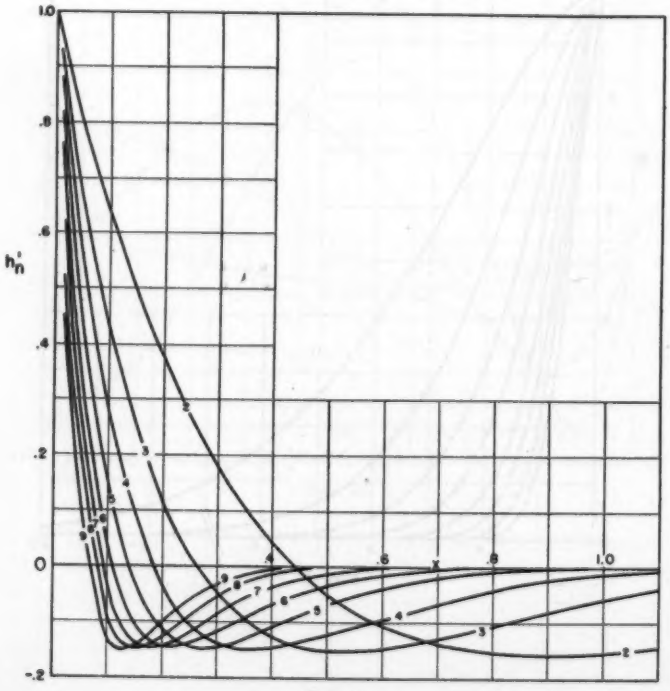

1.0
.8
.6
$h_n'$
.4
.2
0
-.2
.4
.6
X
.8
1.0
2
3
4
5
6
7
8
9

FIG. 6

exact solutions[2-7] indicates that they constitute very good approximations, and are adequate for most engineering purposes.

The functions $f_n(y)$, $f_n'(y)$, $g_n(x)$, $g_n'(x) = -(\alpha_n^2 + \beta_n^2)h_n(x)$, $h_n'(x)$ ($n = 2$ to 9) are plotted in Figs. 2, 3, 4, 5, 6.

General Electric Co., Knolls Atomic Power Laboratory[8]

(Received December 8, 1953)

---

[2] G. Horvay, Discussion of Ref. 1, J. Appl. Mech. *20*, 576 (1953).

[3] G. Horvay and J. S. Born, The Use of Self-Equilibrating Functions in Solution of Beam Problems, Proc. Second Nat'l Congress of Applied Mechanics, in press.

[4] G. Horvay, Thermal Stresses in Rectangular Strips I, Proc. Second Nat'l Congress of Applied Mechanics, in press.

[5] J. S. Born and G. Horvay, Thermal Stresses in Rectangular Strips II, J. Appl. Mech. in press.

[6] G. Horvay and F. N. Spiess, Orthogonal Edge Polynomials in the Solution of Boundary Value Problems, Quarterly of Applied Mathematics *12*, 57 (1954).

[7] G. Horvay and J. S. Born, Precise Closed-Form Solutions of Some Mixed Boundary Value Problems of Plane Elasticity, to be submitted shortly to J. Appl. Mech.

[8] The Knolls Atomic Power Laboratory is operated by the General Electric Company for the Atomic Energy Commission. The work reported here was carried out under contract No. W-31-109 Eng. -52.

# NOTE ON AN INTEGRAL OF RAMANUJAN

By Malcolm K. Brachman

An examination of the inversion of the Laplace transform sech $[\lambda(1 + \mu s)^{1/2}]$ led to consideration of the definite integral

$$I(a, b, c) = \int_{-\infty}^{\infty} \operatorname{sech} axe^{-b(x-c)^2}\, dx. \tag{1}$$

It became evident that information concerning this three-parameter integral could be obtained from a study of the two-parameter integral

$$J(\alpha, \beta) = \int_{-\infty}^{\infty} \operatorname{sech} \pi x \cos \alpha x e^{-\beta x^2}\, dx, \tag{2}$$

since there exists the relationship

$$I(a, b, c) = \sqrt{\pi/b}\, J(2ac, a^2/b). \tag{3}$$

This covers all cases of interest, for it is obvious that $I(a, o, c) = \pi/a$. After completion of the major portion of the current work, a perusal of Titchmarsh's book of Fourier integrals[1] showed that Ramanujan[2] had previously obtained most of my results by treating the integral

$$\phi_w(t) \equiv \int_0^{\infty} \frac{\cos \pi t x}{\cosh \pi x} e^{-\pi w x^2}\, dx = \tfrac{1}{2} J(\pi t, \pi w) \tag{4}$$

It may be noted that $J(\alpha, \beta)$ is essentially the Fourier transform of the product of two functions, both being "self-reciprocal" functions, i.e., functions which are their own Fourier transforms.

This fact negates many of the customary devices for evaluating such integrals. For instance, it precludes using the method given by Titchmarsh for the integral $J(\alpha, i\pi)$. Nevertheless, it proves possible to evaluate integrals of the form $J((2l + 1)\beta + im\pi, \beta)$, where $l$ and $m$ may have any integral values, and many others. A listing is found in the second paper of Ramanujan.

The integral $J(o, \beta)$ is of particular interest. I have not succeeded in evaluating it in closed form except for $\beta = in\pi$, with $n$ or $1/n$ integral. I am indebted to Professor Einar Hille for pointing out to me that $J(o, \beta)$ is the Stieltjes transform of $s^{-1/2}\vartheta_2(\pi, e^{-s})$.

It is possible to derive the following relation:

$$J(\alpha, \beta) + J(\alpha + 2i\pi, \beta) + e^{\alpha+\beta}[J(\alpha + 2\beta, \beta) + J(\alpha + 2\beta + 2i\pi, \beta)] = 0 \tag{5}$$

This suggests a convenient method for summarizing the values of $\alpha$ that allow the integral to be evaluated explicitly. Select a fixed value of $\beta$ such that

$$-\tfrac{1}{2}\pi \leq \arg \beta \leq \tfrac{1}{2}\pi \tag{6}$$

[1] E. C. Titchmarsh, Introduction to the Theory of Fourier Integrals, Oxford, 1937.

[2] S. Ramanujan, Messenger of Math. **44** 75 (1915); Journal Indian Math. Soc. **11** 81 (1919).

Consider $\alpha$'s of the form

$$\alpha = l\beta + mi\pi \tag{7}$$

Take $\beta$ and $i\pi$ as orthogonal unit vectors of a two-dimensional coordinate system. It is then necessary to consider only non-negative values of $l$ and $m$, i.e., the first quadrant of our coordinate system. It turns out that it is possible to evaluate $J(\alpha, \beta)$ for the $\{l, m\}$ values $\{1, 0\}$, $\{0, 1\}$, and $\{1, 1\}$ for arbitrary $\beta$ satisfying (6), and for any additional lattice points which can be reached from these by adding arbitrary integral multiples of the vectors $\{2, 0\}$ and $\{0, 2\}$. In the degenerate case $\beta = ni\pi$, $n$ or $1/n$ integral, we may evaluate $J(\alpha, \beta)$ at the additional point $\{0, 0\}$ and thus at all lattice points $\{l, m\}$, $l$ and $m$ integral.

TEXAS INSTRUMENTS INCORPORATED
DALLAS, TEXAS

(Received March 15, 1954)